

The Journal of Abnormal and Social Psychology

Editor: M. BREWSTER SMITH, *New York University*

Associate Editor: E. J. SHOEN, JR., *Teachers College, Columbia University*

CONSULTING EDITORS

GORDON W. ALLPORT
Harvard University

URIE BRONFENBRENNER
Cornell University

ROGER BROWN
*Massachusetts Institute of
Technology*

DONALD T. CAMPBELL
Northwestern University

LAUNOR F. CARTER
*System Development Corpora-
tion*

RICHARD CHRISTIE
Columbia University

MORTON DEUTSCH
Bell Telephone Laboratories

I. E. FARBER
State University of Iowa

DONALD W. FISKE
University of Chicago

ROGER W. HEYNS
University of Michigan

ROBERT R. HOLT
New York University

IRVING L. JANIS
Yale University

HAROLD E. JONES
University of California

HAROLD H. KELLEY
University of Minnesota

LEONARD S. KOGAN
*Community Service Society
of New York*

DANIEL J. LEVINSON
Harvard Medical School

DAVID C. MCCLELLAND
Harvard University

LEO POSTMAN
University of California

ELIOT H. RODNICK
Duke University

NEVITT SANFORD
University of California

SEYMOUR SARASON
Yale University

RICHARD L. SOLOMON
Harvard University

RENATO TAGIURI
Harvard University

Managing Editor: ARTHUR C. HOFFMAN

Promotion Manager: HELEN ORR

Editorial Assistant: BARBARA CUMMINGS

VOLUME 58
1959



PUBLISHED BY
THE AMERICAN PSYCHOLOGICAL ASSOCIATION, Inc.

MT. ROYAL & GUILFORD AVES., BALTIMORE 2, MD.
AND 1333 SIXTEENTH STREET N.W., WASHINGTON 6, D.C.

CONTENTS OF VOLUME 58

ARTICLES, CASE REPORTS, CRITIQUE AND NOTES

	PAGE
AMEEN, L. See LAFFAL, J.	
AULD, F., JR., & WHITE, ALICE M. Sequential Dependencies in Psychotherapy.....	100
BARNLUND, D. C. A Comparative Study of Individual, Majority and Group Judgment.....	55
BERGEN, B. J. See FREEMAN, H. E.	
BIERI, J., LOBECK, R., & GALINSKY, M. D. A Comparison of Direct, Indirect, and Fantasy Measures of Identification.....	253
BIRNEY, R. C. The Reliability of the Achievement Motive.....	266
BLACKMAN, L. S. See SPITZ, H. H.	
BOOMER, D. S. Subjective Certainty and Resistance to Change.....	323
BOSLEY, J. J. See DESOTO, C. B.	
BREHM, J. W. Increasing Cognitive Dissonance by a Fait Accompli.....	379
BREHM, J. W., & COHEN, A. R. Choice and Chance Relative Deprivation as Determinants of Cognitive Dissonance.....	383
BREHM, J. W., & COHEN, A. R. Re-evaluation of Choice Alternatives as a Function of Their Number and Qualitative Similarity.....	373
BRONSON, WANDA C., KATTEN, EDITH S., & LIVSON, N. Patterns of Authority and Affection in Two Generations.....	143
BROWN, D. R., & DATTA, LOIS-ELLEN. Authoritarianism, Verbal Ability, and Response Set.....	131
CARLSMITH, J. M. See FESTINGER, L.	
COHEN, A. R. See BREHM, J. W.	
COHEN, A. R., TERRY, H. I., & JONES, C. B. Attitudinal Effects of Choice in Exposure to Counter-propaganda.....	388
COHEN, M. See MURRAY, E. J.	
COWEN, E. L., LANDES, J., & SCHAET, D. E. The Effects of Mild Frustration on the Expression of Prejudiced Attitudes.....	33
DATTA, LOIS-ELLEN. See BROWN, D. R.	
DAVIS, J. M. Personality, Perceptual Defense, and Stereoscopic Perception.....	398
DEFLEUR, M. L. See WESTIE, F. R.	
DESOTO, C. B., KUETHE, J. L., & BOSLEY, J. J. A Redefinition of Social Desirability.....	273
DITTES, J. E. Effect of Changes in Self-Esteem Upon Impulsiveness and Deliberation in Making Judgments.....	348
DITTMANN, A. T. See RAUSH, H. L.	
DORIS, J. Test-Anxiety and Blame-Assignment in Grade School Children.....	181
EDUSON, BERNICE T. Structural Analysis of Dreams: Clues to Perceptual Style.....	335
ENGEL, MARY. The Stability of the Self-Concept in Adolescence.....	211
FESTINGER, L., & CARLSMITH, J. M. Cognitive Consequences of Forced Compliance.....	203
FREEMAN, H. E., SIMMONS, O. G., & BERGEN, B. J. Possessiveness as a Characteristic of Mothers of Schizophrenics.....	271
GALINSKY, M. D. See BIERI, J.	
GEWIRTZ, HAVA BONNÉ. Displacement of Preference as a Function of Avoidance-Avoidance Conflict.....	395
GEWIRTZ, HAVA BONNÉ. Generalization of Children's Preferences as a Function of Reinforcement and Task Similarity.....	111
GILL, M. The Present State of Psychoanalytic Theory.....	1
GOLDSTEIN, M. J. The Relationship Between Coping and Avoiding Behavior and Response to Fear-Arousing Propaganda.....	247
GREEN, HELEN B., & KNAPP, R. H. Time Judgment, Aesthetic Preference, and Need for Achievement.....	140
GUZE, S. B., & MENSCH, I. N. An Analysis of Some Features of the Interview With the Interaction Chronograph.....	269
HAIRE, M. See KIRSCHT, J. P.	
HEYNS, R. W. See LANSING, J. B.	
HILLS, J. R. See RAINE, W. J.	
HOFFMAN, L. R. Homogeneity of Member Personality and its Effect on Group Problem-Solving.....	27
JONES, C. B. See COHEN, A. R.	
KAGAN, J., & MOSS, H. A. Stability and Validity of Achievement Fantasy.....	357
KANFER, F. H. Verbal Rate, Content, and Adjustment Ratings in Experimentally Structured Interviews.....	305
KATTEN, EDITH S. See BRONSON, WANDA C.	
KIRSCHT, J. P., LODAHL, T. M., & HAIRE, M. Some Factors in the Selection of Leaders by Members of Small Groups.....	406
KLINE, N. S. See WITTENBORN, J. R.	

	PAGE
KNAPP, R. H. See GREEN, HELEN B.	
KUETHE, J. L. See DESOTO, C. B.	
LAFFAL, J., & AMEEN, L. Hypotheses of Opposite Speech.....	267
LAMBERT, W. W., TRIANDIS, L. M., & WOLF, MARGERY. Some Correlates of Beliefs in the Malevolence and Benevolence of Supernatural Beings: A Cross-Societal Study.....	162
LANDES, J. See COWEN, E. L.	
LANSING, J. B., & HEYNS, R. W. Need Affiliation and Frequency of Four Types of Communication....	365
LAZARUS, R. S. See VOGEL, W.	
LEVENTHAL, D. B., MCGAUGHRAN, L. S., & MORAN, L. J. Multivariable Analysis of the Conceptual Behavior of Schizophrenic and Brain-Damaged Patients.....	84
LEVINSON, D. J. Role, Personality, and Social Structure in the Organizational Setting.....	170
LEVY, L. H., & ORR, T. B. The Social Psychology of Rorschach Validity Research.....	79
LIVSON, N. See BRONSON, WANDA C.	
LOBECK, R. See BIERI, J.	
LODAHL, T. M. See KIRSCHT, J. P.	
LOEVINGER, JANE, & OSSORIO, A. Evaluation of Therapy by Self-Report: A Paradox.....	392
LURIA, ZELLA. A Semantic Analysis of a Normal and a Neurotic Therapy Group.....	216
MCDONALD, J., JR. Personality and Situational Determinants of Conformity.....	241
MCGAUGHRAN, L. S. See LEVENTHAL, D. B.	
MARTIN, B. See PANEK, D. M.	
MENSH, I. N. See GUZE, S. B.	
MILLER, ALICEMARIE. See TALLAND, G. A.	
MITOS, S. B. Semantic Aspects of Prognosis.....	137
MORAN, L. J. See LEVENTHAL, D. B.	
MOSS, H. A. See KAGAN, J.	
MURRAY, E. J., & COHEN, M. Mental Illness, Milieu Therapy, and Social Organization in Ward Groups.....	48
NADLER, E. B. Yielding, Authoritarianism, and Authoritarian Ideology Regarding Groups.....	408
O'CONNOR, N., & RAWNSLEY, K. Two Types of Conditioning in Psychotics and Normals.....	157
ORNE, M. T. The Nature of Hypnosis: Artifact and Essence.....	277
ORR, T. B. See LEVY, L. H.	
OSSORIO, A. See LOEVINGER, JANE.	
PANEK, D. M., & MARTIN, B. The Relationship Between GSR and Speech Disturbances in Psychotherapy.....	402
QUERESHI, M. Y. The Generality of Social Perception Scores.....	317
RACHMAN, S. The Treatment of Anxiety and Phobic Reactions by Systematic Desensitization Psychotherapy.....	259
RAINE, W. J., & HILLS, J. R. Measuring Intra-Individual Variability Within One Testing.....	264
RAUSH, H. L., DITTMANN, A. T., & TAYLOR, T. J. The Interpersonal Behavior of Children in Residential Treatment.....	9
RAVEN, B. H. Social Influence on Opinions and the Communication of Related Content.....	119
RAWNSLEY, K. See O'CONNOR, N.	
RAYMOND, SUSAN. See VOGEL, W.	
RECHTSCHAFFEN, A. See TAYLOR, JANET A.	
SCHAET, D. E. See COWEN, E. L.	
SCHÖNBACH, P. Cognition, Motivation, and Time Perception.....	195
SEGALL, M. H. The Effect of Attitude and Experience on Judgments of Controversial Statements.....	61
SIMMONS, O. G. See FREEMAN, H. E.	
SMITH, E. E. Individual Versus Group Goal Conflict.....	134
SOSKIN, W. F. Influence of Four Types of Data on Diagnostic Conceptualization in Psychological Testing.....	69
SPEERBER, Z. The Test Anxiety Questionnaire: Scoring Norms for a Non-College Population.....	129
SPIZT, H. H., & BLACKMAN, L. S. A Comparison of Mental Retardates and Normals on Visual Figural Aftereffects and Reversible Figures.....	105
TALLAND, G. A., & MILLER, ALICEMARIE. Perceptual Sets in Korsakoff's Psychosis.....	234
TAYLOR, JANET A., & RECHTSCHAFFEN, A. Manifest Anxiety and Reversed Alphabet Printing.....	221
TAYLOR, T. J. See RAUSH, H. L.	
TERRY, H. I. See COHEN, A. R.	
TOMAN, W. A General Formula for the Quantitative Treatment of Human Motivation.....	91
TOWBIN, A. P. Hostility in Rorschach Content and Overt Aggressive Behavior.....	312
TREHUB, A. Ego Disjunction and Psychopathology.....	191
TRIANDIS, L. M. See LAMBERT, W. W.	

	PAGE
VENABLES, P. H. Factors in the Motor Behavior of Functional Psychotics.....	153
VOGEL, W., RAYMOND, SUSAN, & LAZARUS, R. S. Intrinsic Motivation and Psychological Stress.....	225
WEISS, W. The Effects on Opinions of a Change in Scale Judgments.....	329
WESTIE, F. R., & DEFLEUR, M. L. Autonomic Responses and Their Relationship to Race Attitudes.	340
WHITE, ALICE M. See AULD, F., JR.	
WITTENBORN, J. R., & KLINE, N. S. Drive Strength and Symptom Manifestations.....	300
WOLF, MARGERY. See LAMBERT, W. W.	
ZAHN, T. P. Acquired and Symbolic Affective Value as Determinants of Size Estimation in Schizophrenic and Normal Subjects.....	39

THE PRESENT STATE OF PSYCHOANALYTIC THEORY¹

MERTON GILL

Berkeley, California

PSYCHOANALYTIC theory is often regarded by its critics as a rigid Procrustean system. In fact, it is continually changing and growing, although its changes are principally those of widening perspectives, modifications, and additions, rather than alterations in basic assumptions. The past decade has seen a serious effort to make the system of psychoanalytic theory more explicit and internally consistent. Though many people contributed to this effort, I would single out for special mention the work, singly and together, of Hartmann, Kris, and Loewenstein (13, 14, 15, 16, 18, 19, 20, 27, 28, 29), of Rapaport (32, 33, 34, 35, 36, 37, 38, 39), and of Edith Jacobson (23, 24). But it would be incorrect to regard these efforts as simply tidying up the theory. Basic assumptions have been questioned and alternative assumptions have been proposed. A recent paper by Jacobson (24), for example, seriously questions the validity and usefulness of such basic psychoanalytic propositions as primary narcissism and primary masochism. Of course, even more fundamental propositions than these—such as the importance of infantile psychosexuality—have been questioned and discarded by some psychoanalysts (31). The main stream of psychoanalysis, however, holds fast to a number of basic assumptions while adding and modifying others. It is this main stream of the psychoanalytic theory that I will discuss here, though there will be occasional reference to the assumptions of the revisions of psychoanalytic theory. I believe that the ferment in the main stream of psychoanalytic theory is a refutation of the charge that it constitutes an orthodoxy. Any discipline has its share of those who clutch to it as to a religion, and these are not always its strongest members. I will discuss psychoanalytic theory, not psychoanalytic practice. The lag between the two is sometimes great and it may be either theory or practice which forges ahead of the other. It has been gener-

ally true in psychoanalysis that the theory of technique lags behind both basic theory and practice.

IMPORTANT CHANGES IN PSYCHOANALYTIC THEORY

My plan here is to present a number of the basic assumptions of psychoanalytic theory and to indicate what I regard as the important changes in the theory pertaining to each of these assumptions in the last 20 or 25 years.

1. *The Psychoanalytic Theory of Motivation*

Psychoanalysis always has been and continues to be a theory that centers on motivation in human behavior. Uniquely characteristic of psychoanalysis is the kind of motivation it postulates: drives rooted in the biology of the organism. These drives are sexual—in the broader sense in which the word is employed in psychoanalysis—and aggressive. They are characterized by their urgency, their intimate connection with various kinds of bodily behavior, both in terms of one's own body and the bodies of other people, and by the rather bizarre quality of their mode of function when viewed in the light of ordinary conscious motivation.

Psychoanalysis has not given up this conception of primitive drives, but it has somewhat changed its view of their place in personality functioning and has added to its theory to account for other kinds of motivation as well. Whereas the psychoanalytic theory of motivation was formerly restricted almost entirely to primitive drive, now it includes a complex hierarchy of motivations that implies a progressive taming of drives with advancing development and progressive infusion of the drive representations with cognitive elements reflecting external reality (34). Its view of the dynamic relationships of the various levels of the motivational hierarchy has so changed as to increase the emphasis on derivative motivations—to which it has always paid some attention—though by no means diminishing the emphasis on more primitive motives.

¹ Revised from a paper delivered at the AAAS meetings in Berkeley, California, December 1954.

The very term "derivative motivation" implies the earlier psychoanalytic concept of the relationship of various motivational levels. The primitive drives as such were considered to be active even in derivative motivations, which were regarded as reducible to drives at the primitive level. The surgeon was still considered to be expressing his primitive sadistic impulse, the bibliophile to be expressing his anal wishes. The concept which marks a drastic shift from this point of view is that of secondary autonomy (13). Its implications are similar to that of Allport's "functional autonomy" (1) and it conveys the conclusion that derivative motivations develop a semi-autonomy, and can be triggered relatively independently into action. There remains much controversy and unclarity, however, as to the roles played in normal behavior by drives pertaining to the different levels of the motivational hierarchy.

Parallel to the recognition of a motivational hierarchy with degrees of autonomy, there have been developed new concepts of the kind of energy pertaining to motivations. The energy of the primitive drives is libidinal or aggressive but as motivations become progressively autonomous their energy becomes progressively more "neutralized." Such neutralization is said to occur by "delibidinization" and "deaggressivization" (15).

The overthrow of primitive reductionism goes even further, however, than merely to assert that derivative motivations develop a relative autonomy. While formerly all behavior was reduced to primitive drive motivation, now behavior is considered to be determined not only by motivational factors but by other (structural) factors too, which enter into the causal determination of behavior as independent variables and hence are not reducible to motivations, whether primitive or derived.

2. *The Independent Variables Other Than Drive*

Despite the fact that from the very beginnings of psychoanalysis concepts other than motivational were used both in theoretical and clinical studies, the prevailing intent of the theory was to reduce every behavior to motivational terms. A sharp change has come about in that now independent variables

other than motivation are also assumed in the determination of behavior. These independent variables are of two main sorts.

The first are the factors which are conceptualized as primarily autonomous, that is to say, autonomous from drive (13). Whereas secondary autonomy refers to derivatives of drives, primarily autonomous functions arise independently of drives.² These latter are also conceptualized as the apparatuses of the ego and include among others perception, thought, memory, concept formation, and discharge thresholds (34). This is not to say that in any actual perceiving, thinking, remembering, or concept formation, drive factors play no role but rather that independent factors not derivable from drives do play a role in each of these events.

It will be noted that this class of independent variables comprises intraorganismic capacities and structures and that some of them would be called cognitive in general psychological theory. That psychoanalytic theory regards these variables as truly independent is attested to by the relationships conceived between the two kinds of variables—drive and ego apparatus. They are considered to reciprocally influence each other (17) and efforts are made to study both how the development of drive is influenced by the ego apparatuses and how the development of the apparatuses is influenced by drive. It should be noted that the ego apparatuses are only a part of the ego, and the important question has been raised in psychoanalytic theory whether other aspects of ego functioning may not also be primarily autonomous in their origin—as for example, a predilection for the use of a particular variety of defense mechanism (15).

The second class of factors entering into the determination of behavior as independent variables are extraorganismic—those derived from the external environment. These may be classified in various ways—the most usual being a tripartite one into the physical environment, the interpersonal environment, and the social environment.

Naturally both psychoanalytic theory and clinical studies have always dealt with en-

² Whether primarily autonomous functions have available primary neutral energy—that is, neutral energy not derived from libidinal or aggressive energy remains a moot question (16).

vironmental factors, especially the interpersonal ones. The entire field of the development of object relationships—that is to say, relationships with other people—comes under this heading. It is true, however, that the explicit theoretical accounting for the role of the organized social environment is relatively recent, and it is true even that only relatively recently has the effort begun to encompass object relationships systematically within the psychoanalytic theory (27). I realize that this account may be challenged. I am aware that object relationships have long been extensively dealt with in psychoanalytic writings, but I believe that only recently have efforts been made to find a place in systematic theory for the role played by other people in the development of object relationships. In fact I think this lag in psychoanalytic theory proper is one of the reasons that Sullivan's theory (41) has been so favorably received. It made an important contribution in its emphasis on interpersonal relationships but this emphasis became so all consuming as to jettison the psychoanalytic theory of motivation, especially of primitive drive. The same may be said for Horney (22) and with the addition of an emphasis on social factors for Fromm (12).

The problem of the relative roles played by and the relationships between intrapsychic, interpersonal, and social factors is, of course, subject to much controversy. It remains true by and large that the heaviest stress in psychoanalytic theory (insofar as it is to be distinguished from a general psychological theory) will always be on the intrapsychic factors. The outstanding demonstration that there nevertheless is room for social factors within the framework of the established psychoanalytic theory proper is Erikson's work (3, 4, 5).

3. *Maturation in Psychoanalytic Theory*

I will turn now to question whether or not these changes have resulted in any alteration in another one of the basic pillars of psychoanalytic theory, its heavy emphasis on maturational factors in personality development. Psychoanalysis continues to emphasize inborn maturational sequences, but in addition to such sequences in the development of drives, they are now postulated for the development of the ego as well (15). Whatever disputes there may be about the accuracy of the famil-

iar oral-anal-phallic-genital sequence, or the unsettled state of the sequence of stages in ego maturation, psychoanalytic theory continues to put much weight on such maturational sequences. It is true, however, that with the "new regard for the environment" (27), learning has come to occupy a more central role in the psychoanalytic theory of development. There is little agreement in psychoanalytic, just as in general psychological, theory as to the relative roles to be ascribed to learning as against maturation. A temporary re-emphasis of maturation as against learning has taken place in some areas of psychoanalytic theory as a reaction against the postulation of highly complex very early object relationships—as early as the first months of life—by the school around Klein (26). Psychoanalysis has no well-defined and explicit learning theory (21). Yet it seems clear that psychoanalytic theory is incompatible with S-R learning theory and the discussion of its possible relationship to a cognitive or a motivational learning theory like Tolman's (42) is beyond the scope of this paper. It may nevertheless be mentioned, as pointed out by Rapaport (35), that Hartmann's concept of automatization (13) is the psychoanalytic equivalent of habit. Behavior according to this concept is originally learned in a complex motivational context, but may become structuralized, semiautonomous, triggered by external stimuli, and automatized or, in Lewinian (30) terminology, "ossified." The explanation of the development of the motivational life and of interpersonal relationships by maturational rather than learning processes continues to be much more central to psychoanalysis than to general psychology.

4. *Psychoanalysis as a Genetic Theory*

From a discussion of maturation it is natural to turn to another pillar of psychoanalytic theory, its emphasis on longitudinal factors—the genetic point of view. Psychoanalytic theory continues to be heavily genetic in its emphasis. While it is, of course, the case that genetic factors are determinative of current behavior only insofar as they have shaped the form of currently active elements of psychic functioning, it is nevertheless true that psychoanalytic theory holds that many behavioral manifestations become understand-

able only in terms of their genesis. I should mention here two recent major assumptions in psychoanalytic theory that pertain to what we might call general laws of genesis. The first is that while psychoanalytic theory has long held that the phases of ego development occur in close connection with the correlated phases of libidinal development, it is now proposed that the vicissitudes of aggressive drives and of independent factors in the ego must also be taken into account. An example is Hartmann's proposal that "if the defensive reaction against danger from within is modelled after the one to danger from without, it is possible that there too [in the defense mechanisms] the use of—in this case more or less neutralized—aggressive energy is more regular than the use of desexualized libido" (15). The second major suggestion is that primitive id functions may be "taken over" by the ego and form the basis for quite different functions or different uses of the same function. The defensive functions of the ego again offer the most ready examples. As Hartmann (15) says, "introjection . . . probably exists as a form of instinct gratification before it is used in the service of defense. . . . The ego can use, for defense, characteristics of the primary process, as in displacement. . . . Freud [10] has drawn a parallel between the mechanism of isolation and the normal process of attention." A proper understanding of such an altered function would necessarily involve a knowledge of its genesis. The importance of genetic concepts becomes especially obvious in the partial—though to be sure highly modified—return to earlier forms of function in regressive states. In regression there is to a greater or lesser extent a loss of relative autonomy and a reversion of neutralized energy to more aggressive and libidinal forms.

5. *Primary and Secondary Processes in Psychoanalysis*

An exceedingly important foundation of psychoanalytic theory—regarded by some as Freud's greatest achievement—and one as yet little regarded by general psychology is the description of two modes of functioning of the psychic apparatus: the primary and secondary processes (8). The primary process abides by the pleasure principle, and employs the mechanisms of displacement,

condensation, and symbolization. It is a kind of short circuiting of gratification, disregarding the laws of logic and operating with little reference to the nature of external reality. The secondary process is a mode of psychic functioning that abides by the reality principle, by the laws of logic, and takes into account the nature of real external reality. One of the important developments in psychoanalytic theory is that these two modes of psychic functioning are no longer regarded as a dichotomy and as mutually exclusive. They are now conceived as ideal poles of a continuum (28), and an important new concept, proposed by Kris (29), gives account of the fact of the adaptive use of primary process functioning by the ego, namely: "regression in the service of the ego."

It may be worth pointing out here that the shift in the psychoanalytic view of motivation described earlier can from another point of view be described as a shift from emphasis on primary process functioning to secondary process functioning. The mode of operation of primitive drives in the immature organism is largely by primary process mechanisms. In fact it was an unwarranted generalization of this revolutionary discovery that initially led psychoanalysis to subordinate all else to primitive motivational factors. The recognition of derivative motivations and of independent variables other than drives operating in the determination of behavior—more specifically independent variables relating to the nature of the external world—is a shift to the view that much of normal functioning is secondary process functioning.

6. *The Unconscious in Psychoanalysis*

Psychoanalysis continues to insist on the overwhelming importance of unconscious processes in psychological life. It must be pointed out, however, that the significance of unconscious factors is closely related to the psychoanalytic theory of motivation, and alteration in the motivational view will have serious repercussions on the theory of the role of the unconscious. If a good deal of normal behavior is ascribed to derivative semi-autonomous motivations which can be conscious, by just that much is unconscious motivation shorn of its significance in normal behavior. Yet there has not been any real

diminution in psychoanalysis of emphasis on unconscious motivation. In my opinion, this seeming paradox may be at least partially resolved by the view that the semiautonomous motivations are indeed only semiautonomous and that any particular behavior may be viewed in a "nest" of motivational contexts of increasing generality. It is difficult to describe this conception in just a few words, but it argues that a semiautonomous conscious motive operates in a wider context as a means in relation to a less conscious and more primitive motivation, which in turn operates in a wider context as a means in relation to an even more primitive and unconscious motivation. I would like to repeat that the role that a theoretical system is likely to ascribe to unconscious processes will be intimately linked with its view of the motivational structure of personality.

7. *The Introduction of the Structural Point of View*

I can summarize a number of the important changes I have sketched in psychoanalytic theory by uniting them under the heading: the introduction of the structural point of view. It is often said that the structural point of view was introduced into psychoanalysis in 1923 with the publication of Freud's work *The Ego and the Id* (9). In some ways this statement is true but in others false. It overstates the case in that structural considerations are both implicit and explicit in psychoanalytic theory from the very beginning. The concept *topographic* rather than structural was first used and the topographic divisions were made in the dimension of consciousness—the unconscious, the preconscious, and the conscious. Under the impact of the realization that both that which was repressed and the repressing forces too could be unconscious, Freud (9) was led to hypothesize the structures id and ego, and later the topographic divisions of consciousness, hitherto called "systems," became regarded as only qualities rather than coherent systems (11). As long ago as in the seventh chapter of *The Interpretation of Dreams*, consciousness itself was regarded as a supraordinate apparatus. In present theory, this apparatus is ascribed to the ego, yet this early concept of consciousness remains the most sophisticated psychoanalytic view of

its functioning (38). The statement is further false in the sense that the id-ego-superego trichotomy is only a very gross statement of the structural point of view and actually gave names to structures which had, though not systematically, in part been already marked out. Only in the last several decades has the structural point of view been developed in a thorough and systematic manner.

It is not easy to give a clear yet brief picture of this point of view. A full description of a mental process—a metapsychological description—requires its discussion from three points of view, dynamic, economic, and structural. Of these points of view, the dynamic refers to the interplay of forces, the economic refers to energy considerations and to the restoration of the homeostatic equilibrium, while the structural refers to that aspect of the mental process which characterizes its place in the enduring, stable organization of the mind (39). In roughest form, a process is characterized structurally as to its position in one of the three mental structures—id, ego, or superego. Far more subtle dissection of the structure of mind has been carried out as far as ego functioning is concerned. Structural considerations within one of the three major structures are called intrasystemic (15). The description of the ego, for example, as an organization of semiautonomous behavioral dispositions is a structural statement. Early psychoanalytic theory may be characterized as dynamic and economic. If the early psychoanalytic view of motivation had been taken seriously, the most ordinary item of behavior could be conceived as occurring only as a result of the interplay of powerful forces—what Rapaport called "a battle of Titans" (34). The introduction of the structural point of view makes possible a view of personality functioning which includes its steady, stable, ordinary, organized, enduring patterns of behavior and thinking.

These advances in the structural conception are another way of stating the change in the psychoanalytic view of motivation, of energy pertaining to motivation, and even of the recognition of independent variables other than motivation determining behavior, since we may restate these propositions as: derivative motivations arise from relatively autonomous (structuralized) apparatuses employing neu-

tralized energy; the primary autonomous apparatuses are ego structures.³ The ego as a whole is itself conceptualized as relatively autonomous, from the id on the one hand and the environment on the other (37).

ALTERNATIVE FORMULATIONS OF TRENDS IN PSYCHOANALYTIC THEORY

I have now stated in brief the major assumptions of and major changes in psychoanalytic theory. To achieve further clarification I will now describe various other ways in which these same changes have been stated.

One general way of stating the change is that psychoanalysis has moved into the area of ego psychology in comparison with its former preoccupation with id psychology. Ego psychology includes considerations of the autonomous apparatuses, the hierarchy of motivations which are the systems of disposition to behavior as we find them in the ego, and the extra-organismic variables which enter into the determination of behavior by way of the cognitive functions of the ego.

Another way of stating the change is to say that psychoanalysis is now not only a psychology of the depths, but of the surface too (6). The psychological surface is the ego, and it is the ego which is in contact with the external real world.

Yet another way of saying the same thing is that psychoanalysis has included adaptation among its basic concepts (13), whereas formerly it was preoccupied with intrapsychic processes. Adaptation implies an ego in contact with the real external world and the determination of behavior by variables other than motivational. The same principle has been expressed as the increased attention paid by psychoanalysis to the environment (27). I have made this point in another way in my discussion of the primary and secondary processes. The secondary process abides by the reality principle, which is the principle of adaptation. The most general way to state the change: psychoanalysis is becoming a complete psychological system, embracing all of human psychological functioning (13). While psychoanalysis was principally con-

cerned with psychopathology, large areas of personality functioning remained unaccounted for by its theory. As a total psychological system, psychoanalysis had to complement its motivational considerations with cognitive and adaptive ones. The social structure, for example, which at least in some ways plays a more obvious though perhaps no more prominent role in "normal" than in "neurotic" function begins to loom larger in present day psychoanalytic theory.

CHANGING RELATIONSHIPS WITH PSYCHOLOGY

Now a few comments on the changing relationships between psychoanalysis and academic psychology. I have said that psychoanalysis aspires towards becoming a total psychological system by including cognitive and adaptive considerations as well as motivational. It will be well to stress that in part the stimulus for this move comes from academic psychology, though references in psychoanalytic theory to the contributions of academic psychologists are few and far between. It is obvious, though, that cognitive and adaptive considerations have been the special provinces of academic psychology, and just as psychoanalysis has moved to include cognitive functions and to recognize more explicitly adaptive considerations so has psychology moved to a greater emphasis on motivation. The so-called "new look" (2) in research on perception—that is, cognition—is essentially the study of perception in a motivational framework. There is a good deal of thought in general psychology about the relationship of the cognitive and motivational factors. For instance, Klein and Krech (25) consider the dichotomy of cognition and motivation misleading and unnecessary and regard all behavior as consisting of unities which involve cognition and motivation indivisibly. I agree with this point of view and have separated motivation and cognition in this paper only for convenience of exposition.

It must not be presumed, however, that psychoanalysis and psychology are now merging or becoming indistinguishable. Indeed, many people will find it ridiculous that I sound this warning. Nevertheless, it is worth pointing out that once propositions from these perspectives are stated on a sufficiently high level of abstraction, vital differences begin

³ Whether derivative motivations "arise" in a hierarchy from hierarchically ordered means-structures or whether the motivations themselves must be considered structures requires further clarification.

to be washed away. This is most strikingly true with regard to the problem of motivation. On the one hand, both psychoanalysis and psychology stress motivation and both accept the ideas of primary, derived, and relatively autonomous motivations; on the other, the specific picture of the motivational structure of man is vastly different in the two.

Psychoanalysis still adheres to a view of motivation as built on drives rooted in the biology of the organism. However much derived motivations are recognized, psychoanalytic theory still views behavior as essentially motivated by and occurring in the context of bodily drives. Drives for security, success, prestige, status, seem, relative to these basic drives, "superficial" to the psychoanalyst. He thinks rather of castration fear, oedipal wishes, cannibalistic impulses, homosexuality, or a drive to rend and destroy. It is often said that the analyst's preoccupation with these primitive impulses stems from his almost exclusive concern with disturbed personalities. But the analyst believes that the normal person, too, is occupied with dealing with such impulses. The difference between the normal and the neurotic he sees not in the root of the tree of the motivational hierarchy but much closer to its crown.

A word about the libido theory which is tenaciously defended by psychoanalysts because it represents for them the view of motivation described in this paper as classically and specifically psychoanalytic. In some sense this is unfortunate, because as a consequence criticism of the libido theory and demonstration of errors in it are viewed by some groups of analysts as well as by psychologists as an overthrow of the entire psychoanalytic theory of motivation. I would like to suggest that what is really being defended by analysts is not the libido theory as such but the overwhelming importance of relatively primitive bodily drives in motivating behavior. Discussions of the libido theory (40) in my opinion would be more fruitful if it were not regarded as equivalent to the more general psychoanalytic theory of motivation. The place of instinctual libidinal impulses in personality functioning does not stand or fall with the libido theory per se.

Psychologists are also beginning to talk more frequently of unconscious processes in human beings. But here, too, the apparent

rapprochement should be taken with a grain of salt. It is easy to employ the concept of unconscious processes without taking them seriously, while they are taken very seriously indeed in psychoanalytic theory and practice. Behavior, as described earlier in this paper is regarded by psychoanalysis as essentially motivated by unconscious processes. To some extent the difference between the psychoanalyst and the psychologist is the molar-molecular problem. The psychologist is ordinarily concerned with small segments of behavior such as can be observed and controlled in the laboratory, and which do not readily show the presence of unconscious forces in the psychoanalyst's sense. The psychoanalyst on the other hand is more concerned with "real life" behavior and major trends and crises. Some integration of the two points of view may be brought about by what I described earlier as the possibility of viewing any particular behavior as occurring in a "nest" of motivational contexts of increasing generality.

There is much work in contemporary clinical psychology which deals with fairly direct derivatives of unconscious processes; for example, in some types of Rorschach responses. But these must not be confused with the unconscious processes themselves; and just as important is the fact that although the psychologist may recognize these as derivatives of unconscious processes, this should not be taken to mean that the person who has produced these responses is thereby brought any closer to an awareness of, or insight into, the contents of his unconscious.

SUMMARY

To summarize briefly: I have described some of the basic assumptions of psychoanalytic theory and the recent changes which have taken place in them under the headings: 1. The motivational theory; 2. the introduction of independent variables other than drives; 3. emphasis on maturation; 4. the genetic emphasis; 5. the primary and secondary processes; 6. the unconscious; and 7. the introduction of the structural point of view. I then indicated that these changes in psychoanalytic theory are often described as psychoanalysis becoming concerned with: (a) ego psychology; (b) the psychology of the surface; (c) increased attention to the environment; and (d) the introduction of cognitive and

adaptive considerations. I closed with a few general comments on the relationship of psychoanalysis and psychology.

REFERENCES

1. ALLPORT, G. *The nature of personality: Selected papers*. Cambridge: Addison-Wesley, 1950.
2. BLAKE, R., & RAMSEY, G. (Ed.) *Perception, an approach to personality*. New York: Ronald, 1951.
3. ERIKSON, E. *Childhood and society*. New York: Norton, 1950.
4. ERIKSON, E. The dream specimen of psychoanalysis. *J. Amer. psychoanal. Ass.*, 1954, **2**, 5-56.
5. ERIKSON, E. The problem of ego identity. *J. Amer. psychoanal. Ass.*, 1956, **4**, 56-121.
6. FRENKEL-BRUNSWIK, ELSE. Psychoanalysis and the unity of science. *Proc. Amer. Acad. Arts Sci.*, 1954, **80**, 271-350.
7. FREUD, S. *The interpretation of dreams*. Standard Edition, Vols. IV & V. (Originally published 1900) London: Hogarth, 1953.
8. FREUD, S. The unconscious. In *The Standard Edition*, Vol. XIV. (Originally published 1915) London: Hogarth, 1957. Pp. 159-208.
9. FREUD, S. *The ego and the id*. (Originally published 1923) London: Hogarth 1947.
10. FREUD, S. *The problem of anxiety*. (Originally published 1926) New York: Psychoanalytic Quarterly Press, 1936.
11. FREUD, S. *An outline of psychoanalysis*. (Originally published 1938) New York: Norton, 1949.
12. FROMM, E. *Man for himself*. New York: Rinehart, 1947.
13. HARTMANN, H. Ego psychology and the problem of adaptation (abridged). (Originally published 1939) In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951. Pp. 362-396.
14. HARTMANN, H. On rational and irrational action. In G. Roheim (Ed.), *Psychoanalysis and the social sciences*. New York: International Universities Press, 1947. Pp. 359-392.
15. HARTMANN, H. Comments on the psychoanalytic theory of the ego. In *The psychoanalytic study of the child*, Vol. V. New York: International Universities Press, 1950. Pp. 74-96.
16. HARTMANN, H. Notes on the theory of sublimation. In *The psychoanalytic study of the child*, Vol. X. New York: International Universities Press, 1955. Pp. 9-29.
17. HARTMANN, H., HOFFER, W., & FREUD, A. Contributions to symposium on the mutual influences in the development of the ego and id. In *The psychoanalytic study of the child*, Vol. VII. New York: International Universities Press, 1952. Pp. 9-50.
18. HARTMANN, H., & KRIS, E. The genetic approach in psychoanalysis. In *The psychoanalytic study of the child*, Vol. I. New York: International Universities Press, 1945. Pp. 11-29.
19. HARTMANN, H., KRIS, E., & LOEWENSTEIN, R. Comments on the formation of psychic structure. In *The psychoanalytic study of the child*, Vol. II. New York: International Universities Press, 1946. Pp. 11-38.
20. HARTMANN, H., KRIS, E., & LOEWENSTEIN, R. Notes on the theory of aggression. In *The psychoanalytic study of the child*, Vol. III/IV. New York: International Universities Press, 1949. Pp. 9-36.
21. HILGARD, E. *Theories of learning*. New York: Appleton-Century-Crofts, 1956.
22. HORNEY, KAREN. *New ways in psychoanalysis*. New York: Norton, 1939.
23. JACOBSON, EDITH. The affects and their pleasure-unpleasure qualities in relation to the psychic discharge processes. In R. Loewenstein (Ed.), *Drives, affects, behavior*. New York: International Universities Press, 1953. Pp. 38-66.
24. JACOBSON, EDITH. The self and the object world. In *The psychoanalytic study of the child*, Vol. IX. New York: International Universities Press, 1954. Pp. 75-127.
25. KLEIN, G., & KRECH, D. The problem of personality and its theory. In D. Krech & G. Klein (Eds.), *Theoretical models and personality theory*. Durham: Duke Univer. Press, 1952. Pp. 2-23.
26. KLEIN, MELANIE. *Contributions to psychoanalysis, 1921-1945*. London: Hogarth, 1948.
27. KRIS, E. Notes on the development and on some current problems of psychoanalytic child psychology. In *The psychoanalytic study of the child*, Vol. V. New York: International Universities Press, 1950. Pp. 24-46.
28. KRIS, E. On preconscious mental processes. *Psychoanal. Quart.*, 1950, **19**, 540-560.
29. KRIS, E. *Psychoanalytic explorations in art*. New York: International Universities Press, 1952.
30. LEWIN, K. Intention, will and need. (Originally published 1926) In D. Rapaport (Ed.), *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951. Pp. 95-153.
31. MUNROE, RUTH. *Schools of psychoanalytic thought*. New York: Dryden, 1955.
32. RAPAPORT, D. On the psychoanalytic theory of thinking. *Int. J. Psychoanal.*, 1950, **31**, 161-170.
33. RAPAPORT, D. The autonomy of the ego. *Bull. Menn. Clin.*, 1951, **15**, 113-123.
34. RAPAPORT, D. The conceptual model of psychoanalysis. *J. Pers.*, 1951, **20**, 56-81.
35. RAPAPORT, D. (Ed.) *Organization and pathology of thought*. New York: Columbia Univer. Press, 1951.
36. RAPAPORT, D. On the psychoanalytic theory of affects. *Int. J. Psychoanal.*, 1953, **34**, 177-198.
37. RAPAPORT, D. The theory of ego autonomy: A generalization. *Bull. Menn. Clin.*, 1958, **22**, 13-35.
38. RAPAPORT, D. Cognitive structures. In *Contemporary approaches to cognition*. Cambridge: Harvard Univer. Press, 1957.
39. RAPAPORT, D. The structure of psychoanalytic theory (a systematizing attempt). In S. Koch (Ed.), *Systematic resources of psychology*, in press.
40. REIDER, N. (Reporter) Re-evaluation of the libido theory. *J. Amer. psychoanal. Ass.*, 1955, **3**, 299-308.
41. SULLIVAN, H. *The interpersonal theory of psychiatry*. New York: Norton, 1953.
42. TOLMAN, E. A psychological model. In T. Parsons & E. Shils (Eds.), *Toward a general theory of action*. Cambridge: Harvard Univer. Press, 1952.

Received October 28, 1957.

THE INTERPERSONAL BEHAVIOR OF CHILDREN IN RESIDENTIAL TREATMENT¹

HAROLD L. RAUSH, ALLEN T. DITTMANN, AND THADDEUS J. TAYLOR

National Institute of Mental Health

THIS REPORT is concerned with the interpersonal behavior of a small group of disturbed children and with changes in their behavior over a period of a year and a half in a residential treatment program. In general, the study was exploratory, oriented toward description and a search for order in complex behavioral events, toward the evaluation of behavioral change in treatment, and toward the evaluation of a method for observing and coding interpersonal behavior.

Toward Description

The source of data was spontaneous behavior as it appeared in the daily activities of six "hyperaggressive" or "acting-out" boys. To some extent, then, the paper presents a naturalistic study of the social behavior of a small group of children. Studies of small groups have mostly centered on the social interactions of adults in task-oriented situations (Bales, 1950; Hare, Borgatta, & Bales, 1955), whereas interest in the social behavior of children has generally focused on specific variables in relation to more or less specific hypotheses (Baldwin, 1955). The work of Barker and Wright (1954) on the ecology of children's behavior in a small Midwestern town is an exception to the trend. The methods of the present study differ from those of Barker and Wright—a difference in part guided by doubts about the usefulness of a level of approach that is so strictly phenomenological. But there is a major shared value orientation: that it would be useful to have objective and manageable descriptions of the quality, frequency, and intensity of everyday behavior for all sorts of groups and for all sorts of situations. Such a statement is

not an exhortation to random empiricism. Description is not necessarily atheoretical, and an interest in descriptive patterns of everyday behavior is not necessarily antithetical to the testing of hypotheses about specific variables. To some extent, there is, however, a question of order. Anthropologists, for example, have learned that one cannot legitimately interpret specific actions in isolation from other actions or in isolation from patterns within a culture, and they have learned how misleading isolated hypotheses may prove when contextual information is lacking. Adequate psychological descriptions of the everyday behavior of various subgroups, theoretically or pragmatically defined, would constitute a step toward the formulation of specific hypotheses. Such descriptions would also constitute a step toward a general ethology of human behavior.

The group under study was small, and it was a rather special one in the sense that it was selected for a particular syndrome to which considerable social interest accrues in relation to delinquency and which is recognized as difficult to treat clinically. As a group, it was not task-oriented, nor had it come together through the common interests of its members. The environment was also a special one—even if only because psychiatric institutions are not typical of children's living arrangements. Because of the selection process, the children were rather homogeneous, and they lived together under the protection, supervision, and control of a professional staff in the rather homogeneous environment of the institution. The limitations of the present study are, however, perhaps not so much in the special characteristics of the subjects or of their environment, but rather in the lack of comparative data.

Toward Evaluation of Change

The second aim is toward the evaluation of treatment change, not as measured by ratings or by tests, but rather as manifested in the interactions of everyday life. It was, of course, expected that the children would show some

¹ Among the people, other than the authors, who have been involved in this study at one time or another, special note should be made of the contributions of Donald S. Boomer and D. Wells Goodrich. They were not only contributors to the observation and coding processes, but the procedures and methodology used here owe much to them. Fritz Redl, Chief of the Child Research Branch, and the Child Care Staff made the study possible by their cooperation.

improvement in their overt interpersonal relations. Findings of improvement would be consistent with clinical impressions about most of the children. But long-term clinical work has its hazards, one of which is that changes, if and when they occur, are often so gradual that an awareness of the similarities and contrasts between the then and the now is easily lost or distorted. A more formal method would exercise a discipline on both clinical wishes and clinical doubts, while perhaps also pointing to areas and problems which were less clinically obvious.

Clearly, the study is not definitive in relation to treatment of the "acting-out" child or in relation to any of the specific variables that enter into the residential treatment situation. Critical tests of treatment benefits can most relevantly be made outside the framework of a therapeutically focused environment; critical tests would involve variables in addition to those overt aspects of interpersonal behavior which are investigated here; and finally, critical tests would involve adequate controls. What can be fruitfully investigated is whether a group of children change systematically in their ordinary interactions. Given evidence of such changes, examination may be made of their relevance to treatment aims. From the viewpoint of therapeutic concerns, the possibilities of favorable modification in the behavior of a clinically difficult subgroup might be shown, and perhaps the clinical worker may be encouraged in his efforts.

Toward Evaluation of the Method

At the least, a method for evaluating interpersonal behavior and behavioral change was studied. The method involved multiple observations of children in naturalistic settings and the coding of these observations by a scheme which was originally devised for studying the group behavior of adults (Freedman, Leary, Ossorio, & Coffey, 1951; Leary, 1957). A partial proof of the method lies in its success or failure in demonstrating expected phenomena. That is, one aspect of the study is related to a question of construct validity (Cronbach & Meehl, 1955). To the extent that the approach achieves the differentiations that might reasonably be expected of it, the method

offers promise for investigations of other groups in other environments.

METHOD

The Children and the Institution

The six boys were the total patient population of a hospital ward which they entered when they were from 8 to about 10 years old. There was—and still is—no diagnostic category into which they fit easily. In general, their pathology and their actions were beyond the realm of the typical childhood neuroses, yet at the same time they did not represent childhood psychoses. The behavior of the children was characterized by such overwhelming aggressiveness that they could not be tolerated by community, schools, foster parents, or parents. Four boys had been referred to courts for destructive behavior, and three of these had been sent to a reformatory shortly before their admission. The two children who had not come to the attention of the courts had been excluded from several schools because of their antisocial actions. There was usually a history of multiple contacts with social agencies, but outpatient clinical treatment and special school programs seemed to be of little use, at least in the long run, in these cases. None of the boys were "gang" delinquents. Their problems seemed rather a function of intense personality disturbances with a marked deficiency in ego controls, particularly where aggression was concerned. They were children such as Redl and Wineman (1957) have described, children often called hyperaggressive or "acting-out," although neither of these terms is ideal. All six were physically healthy, and so far as could be judged from psychological and psychiatric examinations, they were of normal intelligence and showed no evidence of gross brain damage. They came from socioeconomic milieus ranging from lower-lower through lower-middle class, with two children from lower-middle class homes and the rest lower in socioeconomic status.

Throughout the time of these investigations, the children lived on the ward. Their program was planned intensively and minutely. All were seen approximately four hours weekly in psychotherapy; their schooling, which took place adjacent to the ward, was with specially trained and experienced teachers and clinicians; ward programming and the handling of clinical problems of daily living were closely planned and organized, again by people with considerable experience with disturbed children. Most of all, considerable time and effort were devoted to the coordination and integration of the various levels of treatment. This brief description, though markedly oversimplified, is relevant for interpreting the results of the study.

The present report examines the interpersonal behavior of the children at two phases. When the initial series of observations was made, two of the children had been at the institution between three and four months, and the other four had been there nine or ten months. One may suppose, then, that they knew each other fairly well and that they were familiar with the general pattern of living within the ward. Their ages, at the start of the study, ranged from 8 years, 11 months, to 10 years, 11 months, with the median age

at 10 years. The second series of observations was made some 18 months later.²

The Observations

In each phase of the study, each of the six children was observed twice in each of six settings. There were, thus, a total of 144 observations, 72 in each phase. A number of observers were involved, but each single observation was made by a single observer.³ The observer would go over to the ward or to the gym or to the outdoor play area, for example, to observe a particular child according to his assignment of children and settings. He would concentrate on that particular child, trying to follow as much as was possible the transactions between this child and other children or adults. But he would also try to note what went on among the other children within the locus of observation. The observer did not take notes; he would gauge the length of his observation to the amount of the specific activity that he could remember. This meant that the time period of observation was highly variable. If there was, for example, an extremely rapid interplay of behavior, the observation time might be as brief as several minutes; on the other hand, it might extend, if the interchanges were few and far between, to as much as 20 minutes or a half hour. The mean time for an observation was about 8 minutes. An analysis of variance indicated that differences in observation time allotted to either individual children or to different settings were not significant.

After leaving the setting, the observer would immediately dictate onto tapes or Audograph discs a factual, descriptive report of his observation. Observers were cautioned to focus on interaction, to be as specific and concrete as possible, and to avoid psychological terms and inferential conclusions. In general, observer training required only several practice trials.⁴ The protocols of the observations resemble those obtained by Barker and Wright (1954), though their descriptions were undoubtedly more fully detailed.

The Settings

Different situations—for example, even different games—exert a pull for different behaviors, and they differ in the kinds of behavior they sanction, positively encourage, or inhibit (Redl & Wineman, 1957). The opportunity rarely exists to study behavior in a representative variety of settings, to relate behavior to what Brunswik (1947) has called "the ecology of environmental events." A residential treatment center provides such an opportunity. The design of the study utilized a somewhat representative sampling of the kinds of

activities around which the children lived their lives so as to allow investigation of the relevance of different settings for interactive behaviors. This aspect of the study will not be discussed here, but the settings should be noted: (a) breakfast—an early morning observation; (b) snacks in the period just before the children went to bed at night; (c) other mealtimes; (d) structured game activities, ranging from cards to basketball;⁵ (e) unstructured group activities where the specific external task structure was minimal—for example, social conversations; and (f) an arts-and-crafts period. Since one phase of the study occurred during the summer, when school was not in session, this last setting was utilized as an approximation to an instructional situation. Since selection was in terms of kind of activity, the data are not representative of time as a time sample would be; the nature of the settings did, however, insure that various times of the day were included. Relevant settings omitted from the study were psychotherapy, other two-person group situations, where the presence of an observer might be intrusive, and school situations other than arts and crafts.

The Coding Scheme

The search for methods for describing human interactions and the problems involved in extant approaches have been reviewed by Heyns and Lippitt (1954). The present need was for a scheme (a) applicable to a wide variety of situations, (b) relevant for the study of personality and individual behavior patterns, (c) suitable for dealing with the behavior of both children and adults, and (d) relatively comprehensive. The approach initially described by Freedman et al. (1951), and discussed in detail by Leary (1957), was used.

The scheme (Fig. 1) is based on two polar coordinates. One is along the dimension of affection: love (affiliate, act friendly) to hate (attack, act unfriendly). The other axis is concerned with status: dominate (command, high status action) to submit (obey, low status action). Each action of one person toward another is coded by letter into one of 16 categories along the periphery of the circle in accordance with its blending of the two coordinates. The words below the letters are simply examples of the kinds of actions that might be coded at that position. In practice, coding was generally a compromise between the words representing the categories and the position relative to the two axes, but in cases of doubt the position was utilized rather than the words.⁶

As in the Bales scheme (1950), the attitude taken by the coder is that of the "generalized other." The question he attempts to answer via his categorization is, "What is this person doing to the other? What kind of relationship is he attempting to establish through this particular behavior?" (Heyns & Lippitt, 1954, p. 91). For example, when a child says, "Wasn't that a good movie we saw last night?" he is generally not coded

² The first series was completed within a month; the second series took some five months to complete.

³ In addition to the authors and others already mentioned, Joseph H. Handlon and Jeston Hamer served as observers.

⁴ The problems of making formal observations in an on-going clinical operation and the methodological issues in pretesting the approach are complex matters. Discussions of these questions by A. T. Dittmann and by D. W. Goodrich are in preparation.

⁵ Behaviors specific to the game itself—for example, passing a ball, playing a card, or claiming one's turn—were not coded or considered in the analyses to follow.

⁶ Practical coding problems were sometimes resolved by double codings, but such solutions are not wholly adequate.

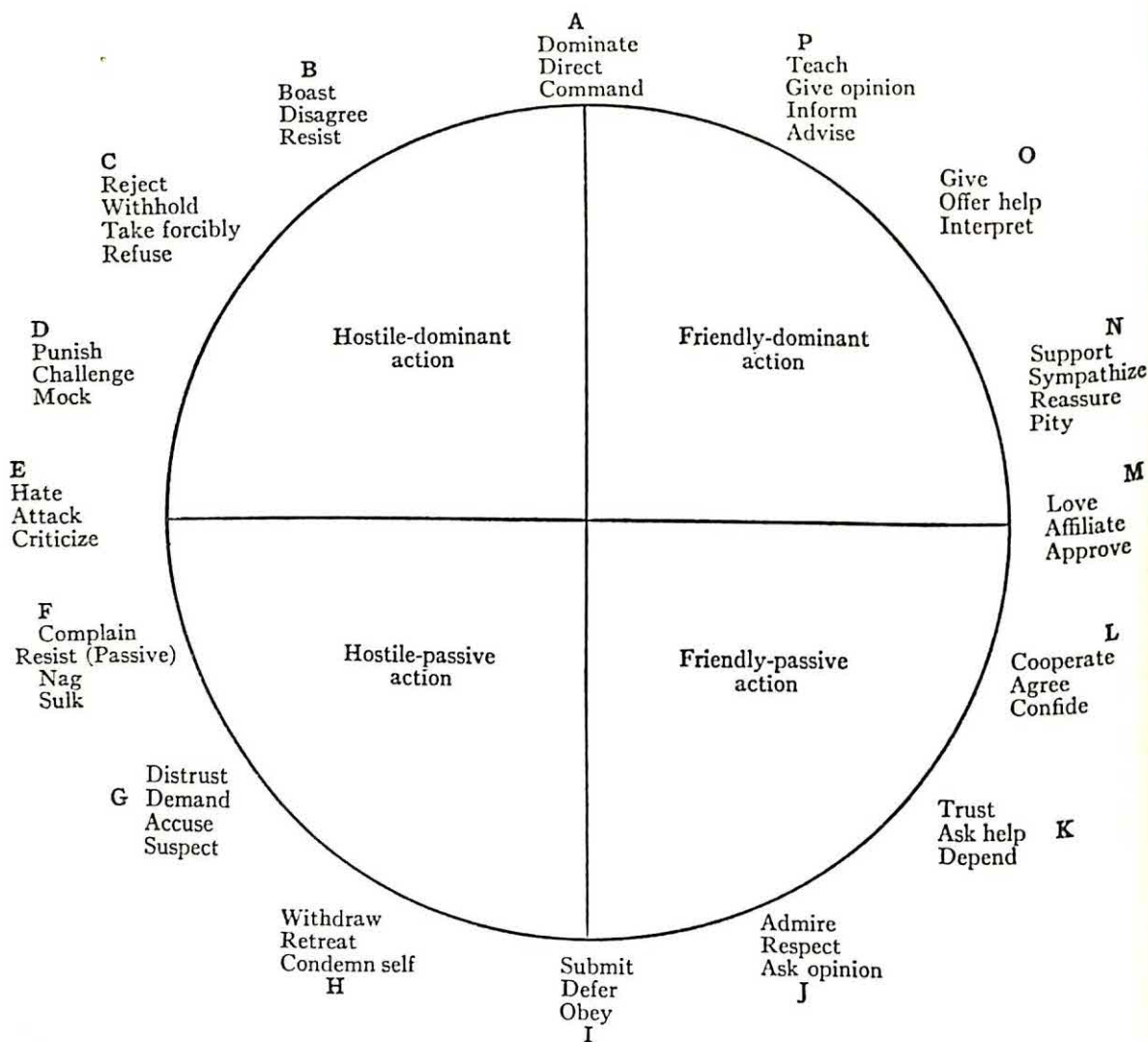


FIG. 1. CATEGORIES OF INTERPERSONAL BEHAVIOR
Modified from Freedman et al. (1951) and Leary (1957, p. 65)

J (Fig. 1), although J can include asking someone's opinion; he is rather coded M, which is simple affiliation. Or when a child says to another, "I can kick better than you," rather than simply stating an opinion P, he is usually establishing a dominant, slightly hostile relationship B, although within some contexts such a statement might involve more of an aggressive element than one of status differentiation and so might be coded D. The statement, "I don't want to play with you," may represent active rejection C, whiny complaining F, or very passive withdrawal H, depending on the context and on the quality with which it was said.

In addition to the categorization described above, a form of "intensity" coding was employed simultaneously. Each interaction was also coded as to whether the behavior was (a) uninvolved—for example, a very casual "Hi" or a casual ignoring of another's statement or request, (b) involved and appropriate, or (c) involved and inappropriate—either an overly intense behavior

or a behavior that is qualitatively inappropriate such as responding to an affectionate gesture with an attack. These latter codings are rather crude, and the aspects of involvedness, appropriateness, and intensity are confounded. Recognizing these limitations, we can still make use of these categories over a wide sampling of behavior, and in point of fact, they do add another useful dimension to our findings.

Coding was done from tapes or typescripts of the dictated observations. Each observation was coded by at least two coders working together. The coders read the protocol (or listened to the tape) line by line, coding each interaction in terms of the person behaving (the specific child or adult), the interpersonal quality and "intensity" of the behavior, and the person toward whom the behavior was directed. Thus, one may obtain in chronological sequence the interactions of any given child toward any other child or adult, and also

the behavior of others, children or adults, towards him.⁷

⁷ Freedman et al. (1951, p. 155) present some data on interrater agreement in coding verbal behavior of adults. Adequate assessment of reliability is made difficult by the fact that there is no baseline for evaluating correlational indices with data scored in this fashion. Rank order correlations between codings from protocols of independent observers of the same events and between pairs of coders of the same protocols are invariably high, but they are likely to be somewhat spurious, since there is some doubt that correlations between random protocols are of zero order. Dittmann (in press) discusses a number of aspects of the reliability of the system. Considering agreement between two pairs of coders working with the same material, Dittmann notes that item-by-item agreement, analogous to test item reliability, is far greater than chance expectancy, but he also notes that there remain appreciable discrepancies. When, however, the single interactions are grouped to form a profile for an observation, a situation analogous to test reliability, discrepancies between independent pairs of coders are far smaller than could be expected by chance. Similarly, the protocols from independent observers of the same events yield smaller differences than could be expected by chance. A recent check by the present authors compared different observers who observed the same children in matched, rather than identical, settings—a situation analogous to alternate forms. Differences, which might have resulted from observer variations or from lack of equivalence in the matched settings or from both, were well within chance limits. Furthermore, a series of observations made approximately two months apart with the same children also failed to yield significant differences.

Clearly, none of these results allows a statement about level of reliability comparable to the usual Pearson r . The significant item-by-item agreement, together with the failure to find evidence of bias in the sources tested, and together with the finding of consistent individual differences, discussed below, would, however, seem to warrant the conclusion reached by Dittmann (1958). The conclusion is that reliability is adequate for grouped data—such as the individual profiles considered in the present study—although it may not be adequate for the analysis of single sequences.

There remains a question of possible bias occurring between the two phases of the study, and the only answers at present are indirect ones. First, the scheme seems fairly objective; second, raters have been aware of the problem, and their continued sensitivity to the possibility of bias has probably served to reduce that possibility; third, the data yield negative as well as positive results, whereas a bias would be likely to operate more consistently. It is recognized that such answers are incomplete. A more definitive check, which has awaited the presence of uncontaminated raters for whom protocols can be adequately disguised, is currently in process.

RESULTS AND DISCUSSION

In the discussion that follows, the basic source of material was, for each child, the 24 protocols—12 in each phase—in which he served as the central focus for observation. For analyzing qualitative changes, the interactions of individual children were distributed into the four quadrants of the circle (Fig. 1); frequencies at the midpoints (M, I, E, and A) were divided evenly between the two adjacent quadrants with any remaining odd entry randomly assigned.⁸ Further data from individual segments of the 16-category scheme are noted occasionally, where this would seem to provide clarification. The approach to analyzing the reaction of others to individual children is commented on later.

The statistical method employed was chi square which, though it fails to take into account the continuity postulated in the circle scheme, involves few assumptions as to the nature of the data. The indices suggested by Leary (1957, pp. 68–71), while mathematically more elegant, would seem to require assumptions even beyond those of ordinal classification. The data for each child were analyzed independently in order to avoid confounding. Where individual chi squares for each child are summed to yield a total estimate of group change, the formula is the sum of chi squares in one direction minus the sum of chi squares in the other direction. Since there are no rational bases for expected distributions of behavior—any assumption that behavior should be distributed equally into each category is obviously untenable—total marginal distributions for individual children were used in obtaining theoretical values. Significance tests are reported for two-tailed distributions, for although some expectations are rather obviously directional in studying behavior change for this group of children, we were at this stage interested in exploring both sides of any coins which seemed worth examining.

⁸ For the early phase observations, the median was 98 interactions for a child toward other children (range 62 to 141), and the median was 89 interactions toward adults (range 66 to 143). In the observations made 18 months later, the median number of interactions toward children was 77 (range 53 to 89); toward adults, it was 69 (range 61 to 98).

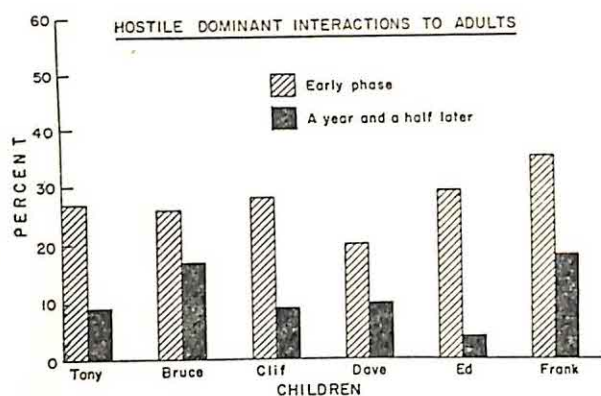


FIG. 2. CHANGES IN PROPORTIONS OF HOSTILE-DOMINANT INTERACTIONS BY CHILDREN TO ADULTS

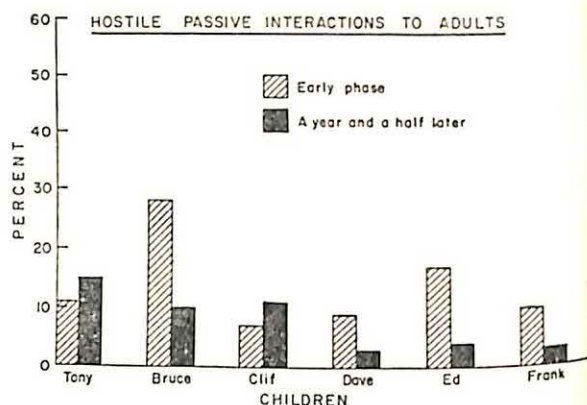


FIG. 3. CHANGES IN PROPORTIONS OF HOSTILE-PASSIVE INTERACTIONS BY CHILDREN TO ADULTS

Changes in Interactions Toward Adults

Hostile-dominant interactions. Figure 2 presents for each of the children the percentage of his total responses toward adults which were coded as hostile-dominant at each of the two phases. The category represents active forms of aggressive behavior. Included are such interactions as actively refusing to comply with adult requests, making boastful demands on adults, threatening or challenging adults, attempting to "argue down" an adult, poking unfriendly fun at an adult, ordering adults around in a boastful or unfriendly manner, and any attack on the adult in his role of authority with the attempt to negate or degrade the authority component. Difficulties in authority relationships were prominent in the case records of all six boys prior to their admission. In one sense, the chief symptoms which entered into their selection were their consistent failure to accept the roles that adults define for children and their active rebellion against adult authority.

In the early phase, the mean percentage of hostile-dominant interactions toward adults was 28. This proportion is, of course, difficult to evaluate without control studies of more normal pre-adolescents in a comparable environment. For the clinical staff, however, it is a patent understatement to say that the amount of active hostility shown by these children was high.⁹ In the later phase, the mean percentage of hostile-dominant responses toward adults dropped to 11. Over the year and a

half between the two phases, each of the children changed in the expected direction. The shift was examined for each child separately by chi square—comparing frequencies of hostile-dominant vs. all other behavior toward adults over the two phases. The changes were significant for one child at $p < .001$, for two children at $p < .01$, for one child at $p < .02$, and for two children at $p < .20$. The sum of the individual chi squares was 44.23, which with 6 *df* (one for each two-by-two table) indicates a change at a level of confidence well beyond .001.

Hostile-passive interactions. Hostile-passive modes of expressing hostility were, in general, less prominent in the relations of the children with adults than were the more dominant forms of aggression discussed above. Such behaviors as whining and complaining in relation to adults, accusing adults of punitive behavior or attitudes, demanding something of an adult in such a way as to imply that the adult is an ungrateful monster, sulky withdrawal from interaction, tearful refusals—these constituted a mean of 14% of the behavior toward adults in the earlier phase (Fig. 3). A year and a half later, the mean percentage had dropped to 8. Four of the six children showed a decrease in passive expressions of aggression—one at $p < .01$, one at $p < .02$, one at $p < .10$, and one at $p < .20$; two children showed an insignificant increase ($p < .50$ in both cases). The sum of the four chi squares in the expected direction minus the two chi squares in the contrary direction is 17.62, which with 6 *df* is significant at a confidence level beyond .01. Thus, it would

⁹ Preliminary data from control studies confirm staff impressions.

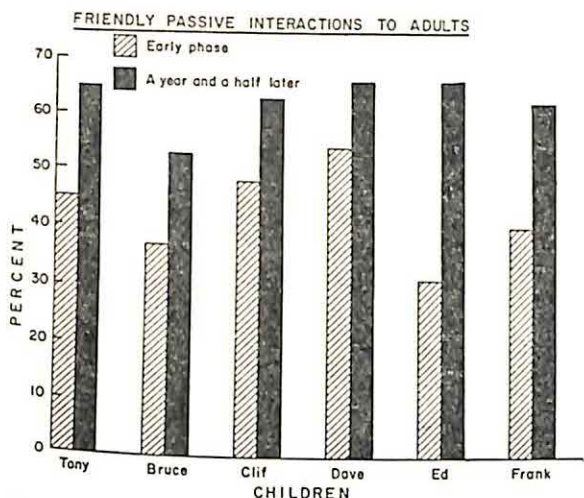


FIG. 4. CHANGES IN PROPORTIONS OF FRIENDLY-PASSIVE INTERACTIONS BY CHILDREN TO ADULTS

seem that while the magnitude of decrease in passively hostile interactions toward adults was significant, the change was not nearly as great as in the case of dominant expressions of hostility.

Friendly-passive interactions. It is interesting that despite the fact that these were hyper-aggressive children, selected because of their unmanageability, the modal response toward adults was in the friendly-passive category (Fig. 4). Even in the earlier phase, friendly-passive behaviors made up the greatest proportion of interactions with adults for each of the children with a mean of 43%. Examination of the total interactions toward adults of all six children when each was in the primary focus of observation indicates that in the earlier phase the three highest ranking of the 16 categories (Fig. 1) were affiliative behaviors (M), cooperative behaviors (L), and help-seeking behaviors (K), which produced respectively 15, 14, and 12% of all interactions toward adults.¹⁰ This phenomenon, rather than indicating that the children were not "really" hyperaggressive, would seem to point in two directions. First, these are children and their behavior in many ways must resemble that of other children of their age. Second, each interaction in these analyses carries a single weight, and while this arrangement serves its purposes, it is unlikely that the recipient, for example, is equally impressed by one friendly

¹⁰ Active resistance or disagreement (B) was next with 10%.

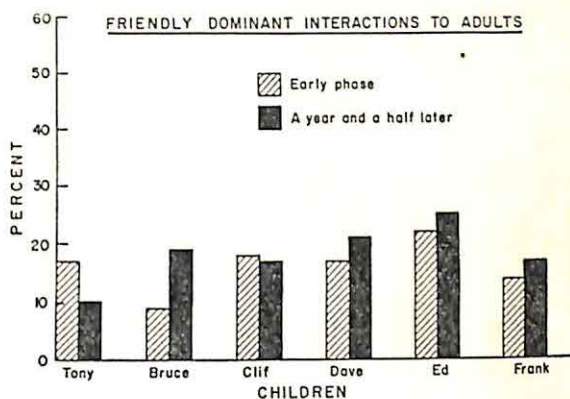


FIG. 5. CHANGES IN PROPORTIONS OF FRIENDLY-DOMINANT INTERACTIONS BY CHILDREN TO ADULTS

"hello" as by a single attack of murderous rage. The relatively high frequencies of friendly-passive behaviors do not negate the difficulties in living with these children. The problem is that there are no units for effectively gauging the psychological impact of an action on the recipient.

By the time of the later period, the percentage of friendly-passive interactions with adults had risen to a mean of 63%. Each of the children showed an increase in such responses—one at $p < .001$, two at $p < .01$, one at $p < .05$, one at $p < .10$, and one at $p < .30$. The sum of the chi squares is 45.37, which with 6 df is significant at a confidence level well beyond .001. Further consideration of the specific nature of the changes appears below.

Friendly-dominant interactions. It is difficult to know what to expect in the friendly-dominant area, both in proportions and in actual changes. In contrast with the situation for aggression and dependency, there has been little theoretical or research interest in such children's behaviors as sympathizing with or reassuring adults, offering help to adults, and teaching or advising adults. One might guess that such actions are perhaps not very appropriate as a major aspect of the relations of pre-adolescent boys with adults. For the present group, friendly-dominant responses constituted a rather small proportion of behaviors toward adults. The values in Fig. 5 and the mean values—16% in the early and 18% in the later phase—are somewhat spuriously high because the major contributory entry was from affiliative responses (M), which in the analyses were distributed equally be-

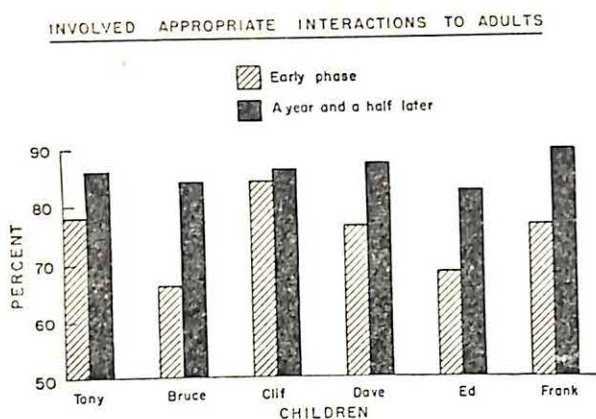


FIG. 6. CHANGES IN PROPORTIONS OF APPROPRIATE-INVOLVED INTERACTIONS BY CHILDREN TO ADULTS

tween friendly-dominant and friendly-passive interactions.¹¹ Only for one child, Bruce, did the change between the two phases approach significance ($p < .10$), and here it was almost wholly a function of the increase in affiliative responses (M). The sum of the four chi squares showing increase minus the two showing decrease was 2.95, which is not significant ($p < .90$).

The "intensity" dimension. Most of the children's behavior was considered by the coders to be appropriate and involved. In the early phase, this "intensity" category comprised a mean of 75% of the children's action toward adults (Fig. 6). Only 9% of the behavior in this period—a mean for the six children—was coded as inappropriate and involved, that is, as being overly intense or qualitatively inappropriate to the circumstances. Uninvolved interactions yielded a mean of 16% of the behaviors in this phase; these included such actions as silent rejections of adult requests, subtle provocations, as well as token gestures of acceptance or affiliation, and the term *uninvolved* is not very adequate.

In any case, the proportions of appropriate-involved interactions increased in the later phase from a mean of 75% to a mean of 86%. Each of the children changed in the expected direction. Comparing the frequencies of appropriate-involved behaviors with the summed frequencies of the other two categories over the two phases yielded chi squares at con-

fidence levels of $p < .01$ for one child, $p < .05$ for two children, $p < .20$ for two children, and $p < .90$ for the sixth child (two-tailed tests). The sum of the chi squares was 21.19, which with 6 *df* is significant at $p < .01$. Five children showed a decrease in inappropriate-involved behavior; the means went from 9% to 4%. All six children showed a decrease in uninvolved actions toward adults, the means dropping from 16% to 10%. The raw frequencies are in some cases too small to warrant statistical analysis, but the trends seem obvious.

Summary comments on changes in interactions with adults. Over the year and a half the children changed considerably in their behavior toward adults. Primarily, they lessened their attempts to dominate adults aggressively, and they increased their friendly and compliant associations with adults. Passive expressions of hostility also decreased, and in general, behavior became more appropriate, but these latter changes, while statistically significant, were less striking.

It is obvious that improvement occurred in overt behavior toward adults. During the period of the study, however, the boys had not only been under an intensive residential treatment program, but they had also grown older. Where so little is known in any systematic way about the interpersonal behavior of normal children and about developmental changes in such behavior, there is the perplexing question that is often put as an issue between maturation and learning (or treatment). The question is a complex one, since social maturation can never be divorced from considerations of the particular environment involved and its indulgences, tolerances, and demands. For example, the treatment of children, in order to be adequate, must take growth and development into account. Conversely, maturational phenomena will be influenced by the environmental medium in which they occur. In the present situation, one may legitimately ask: Are the changes necessarily to be attributed to the treatment program together with growth and development, or are they, perhaps, primarily a function of normal development within a somewhat benign environment? No definitive answer can be given without control studies, and there are manifold problems even so. There are, how-

¹¹ The total of the three categories, N, O, and P, contributed 7% to the total number of interactions in each of the two phases, whereas M contributed 15% and 21% to the early and late phase, respectively.

ever, some partial cues in the data which indicate that the treatment, in a broad sense, is a critical factor. Thus, while one might possibly expect decreased aggression and decreased dominance in relation to adults as part of a "normal" growth process, though the argument would be tenuous, one would not expect to find much increase in trusting, dependent relations with adults. It is interesting that out of the 16 possibilities (Fig. 1), the category that showed the greatest shift was K, which deals with requesting, depending, and asking-help behavior, and that the percentage of responses in this category went from 12% of all behaviors toward adults in the early phase to 24% in the later phase. In frequency of responses, K shifted from third to first rank. It would seem that this was not a maturation phenomenon—that is, it is unlikely that children either become increasingly dependent or increasingly admit their dependency on adults with age. One may speculate that the evidence points to the dissolution of a defensive layer so that dependency emerges; such a speculation dovetails with the impressions of the psychotherapists that oral themes became very prominent in the later phases of psychotherapy with these children. The critical question about what factors in the treatment program contributed to change can unfortunately not be answered by a study such as this, but some further issues in the process of change will be considered below in the discussion of behaviors directed toward each of the children.

Changes in Interactions Toward Peers

We turn to the behavior of the children toward their peers, again considering each child when he was the focal subject for observation, but now investigating changes in the behavior that he directed toward other children. The discussion can be briefer in this section because some of the contextual material has already been presented.

Hostile-dominant interactions. Figure 7 presents for each child at the two phases the percentage of his actions which were hostile-dominant in orientation toward other children. In both phases the means are higher than was the case for behavior toward adults—39% as compared to 28% in the early phase, and 32% as compared to 11% in the later phase. Five

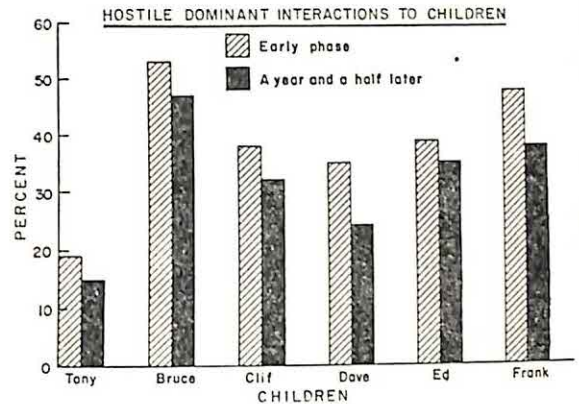


FIG. 7. CHANGES IN PROPORTIONS OF HOSTILE-DOMINANT INTERACTIONS BY CHILDREN TO PEERS

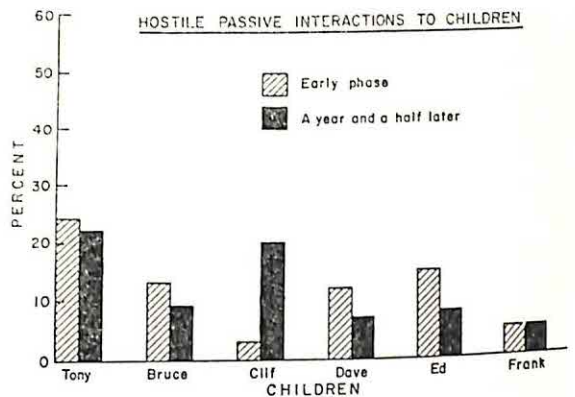


FIG. 8. CHANGES IN PROPORTIONS OF HOSTILE-PASSIVE INTERACTIONS BY CHILDREN TO PEERS

of the six boys were in the early period more dominantly aggressive toward their peers than toward adults, and all six showed this trend in the later period. That is, aggressive attempts at dominance were, as might be expected, more readily expressed toward peers than toward the higher status adults.

All six of the children showed expected decremental changes over the 18 months. While this directional consistency for six cases is significant at a confidence level of $p = .03$ by a simple two-tailed binomial test, none of the individual changes reached statistical significance, and the sum of the individual chi squares was also nonsignificant. Thus, in contrast to the marked shifts between the two phases in proportions of hostile-dominant interactions toward adults, the changes in interactions with peers were, at the most, slight.

Hostile-passive interactions. Although four of the six children showed a decrease in hostile-passive behaviors (Fig. 8), in no case was the

change significant. One child, Clif, increased in passive expressions of aggression toward the others, and the shift was significant by chi square at $p < .001$. For this child, both passive resistance (F) and withdrawal (H) increased over the year and a half. His over-all proportion of hostility rose somewhat, though not significantly, since a slight decline in hostile-dominant interactions did not compensate for the increased passive hostility. The situation thus points to the possibility of a slight deterioration in peer relationships for this child over the period of the study. The mean percentages of hostile-passive responses for the six children are the same in the two phases, 12%, and the sum of the chi squares for the positive change direction minus those in the negative direction is, although opposite to the direction of improvement because of Clif's contribution, not significant.

Friendly-passive interactions. Just as active aggression was more characteristic in response to peers than to adults, so friendly-passive interactions, at the opposite side of the circle, were less characteristic in peer relationships than in relationships with adults. In the early phase, friendly-passive actions made up 43% of the responses to adults and 24% of the responses to other children; in the later period, the respective percentages were 63 and 29. In behavior toward adults, Categories M, L, and K (Fig. 1) held the three highest ranks for response frequencies in each of the two phases, based on the total number of responses of all children in each phase. In behaviors toward peers, M (affiliative responses) and L (cooperative responses) were at Ranks 1 and 4.5 in the

early phase and at Ranks 1 and 2.5 in the later phase. But the position of K (trusting, dependent, help-seeking responses) was quite different in relations with peers. In the early phase, only 2% of all responses were coded K, a ranking of 14 in a triple tie; in the later phase, only 3% of all responses was coded K, a ranking of 11.5 in a quadruple tie. Thus, unlike the case in relation to adults, dependency responses among the children occurred relatively infrequently, and they showed rather little shift between the two times of study.

Four boys showed gains in the proportions of friendly-passive behaviors toward peers (Fig. 9), and for two of them the change was significant by chi square at $p < .05$. None of the other shifts were statistically significant; the sum of the four chi squares in the expected direction minus the two in the opposite direction was 8.87 which, with 6 *df*, yields a confidence level of $p < .20$.

Friendly-dominant interactions. One might expect the friendly-dominant mode of interaction to be more characteristic of relations among children than from children to adults, and such an expectation appears to be legitimate. In the early phase, an average of 26% of the children's responses to peers were friendly-dominant in orientation in contrast to the average of 16% in responses to adults; in the later phase, the two averages were 28% and 18%, respectively. The category M again constituted a major source of data for the quadrant, and that the values are not deceiving is shown by the fact that Categories N, O, and P yielded 15% of all responses toward children

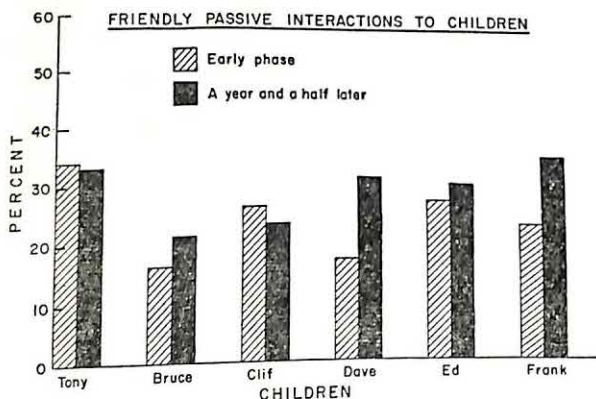


FIG. 9. CHANGES IN PROPORTIONS OF FRIENDLY-PASSIVE INTERACTIONS BY CHILDREN TO PEERS

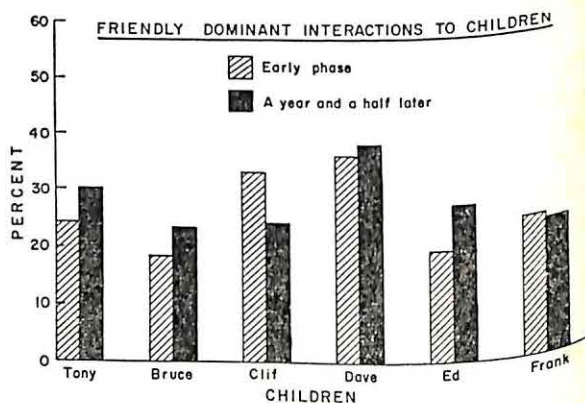


FIG. 10. CHANGES IN PROPORTIONS OF FRIENDLY-DOMINANT INTERACTIONS BY CHILDREN TO PEERS

in each of the two phases and 7% of all responses toward adults in each of the two phases.

Figure 10 shows four of the six changes to be in the direction of an increase in friendly-dominant behaviors toward peers, but none of the changes in either direction is significant; neither is the sum of the chi squares.

The "intensity" dimension. The distributions of responses toward peers are strikingly similar to the distributions of responses toward adults in the three-category "intensity" classification. In the early phase, the mean proportions of interactions toward children were 74% appropriate-involved (as compared to 75% for behavior toward adults), 16% uninvolved interactions (as compared to 16% for behavior toward adults), and 9% inappropriate-involved interactions (as compared to 9% for behavior toward adults). In the later phase, the three means are, respectively, 84%, 11%, and 5% as compared to 86%, 10%, and 4% in behavior toward adults.

Figure 11 shows the shifts in the proportions of appropriate-involved behaviors for each of the children. All changes were in the expected direction of increase. For one child the change was significant at $p < .02$, for another at $p < .05$, and for a third at $p < .10$; for the other three children there is less conclusive evidence of change. The sum of the chi squares is 15.85, which, with 6 df , is significant at a level of $p < .02$. Decreases in inappropriate-involved and in uninvolved actions toward peers seem by inspection to be less consistent than they were in behavior toward adults.

Summary comments on changes in interactions toward peers. Certainly, the evidence of change in relations with peers is much less than in the case of relations with adults. The only general change in peer directed action that warrants much confidence was the increase in the relative proportion of appropriate behavior. In other aspects, the directions were similar to those for behavior toward adults—toward a decrease in dominant aggressive actions and toward an increase in friendly compliant actions—but the shifts over the year and a half were, for the most part, unimpressive.

To the question of why changes in relations to adults were so much more striking than were changes in peer relations, no clear answer can

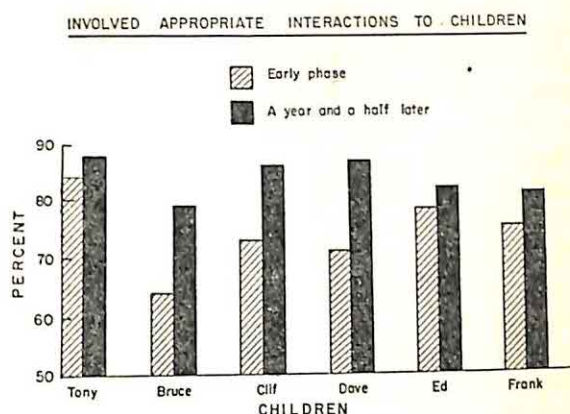


FIG. 11. CHANGES IN PROPORTIONS OF APPROPRIATE-INVOLVED INTERACTIONS BY CHILDREN TO PEERS

be given. There is the possibility that interpersonal behavior with peers was less disturbed than that with adults. Control studies would be required to confirm or disaffirm this statement. There is also the possibility that changes occur earlier in the treatment process in relation to adults than in relation to peers. Such an hypothesis would be reasonable though not necessary on the basis of clinical evidence that the difficulties of these boys developed out of primary relationships with parental figures. Perhaps, with these primitive children, some resolution of earlier relationship problems must occur before the genetically more advanced problems of peer relationships can be met. Follow-up studies and studies of other clinical groups would be useful here. There is also a possibility that the observation method and the instrument are less potent for gauging peer interactions than for interactions of children toward adults. Some light is cast on this latter issue in the discussions that follow.

FURTHER EXPLORATIONS

The aim in the following sections is to explore some phenomena of group and individual interpersonal behavior. Such exploration can perhaps illuminate some of the findings already referred to. It can point to further potentialities or limitations of the methods. Most important, it can probe toward the formulation of general principles.

Group Similarities

The selection of the children and their environmental homogeneity within the institu-

tion should make for behavioral similarities in their interactions. That the children did resemble one another in interpersonal behavior is shown as follows: When the 16 categories were ranked for each child in accordance with the frequencies of his interactions, Kendall W s, computed for the six children and corrected for ties, were .61 for interactions toward peers and .59 for interactions toward adults in the early phase; in the later phase, the corresponding W s were .61 and .76. All W s are significant at levels of $p < .001$. Furthermore, behavior between the two phases correlated significantly. When the interactions for the group taken as a whole in the 16 categories were ranked for each phase, the Spearman rho corrected for ties, for behavior toward peers was .79. Even in the case of interactions toward adults, where the differences between phases were highly significant, the rank orders of interaction categories were similar between the two periods a year and a half apart. The rho, corrected for ties, was .75.¹² Thus, although interpersonal behavior may change over time, it also seems to maintain a certain consistency.

The concordances among individuals and the consistencies in group behavior raise a number of questions. We do not know to what extent correspondences are a function of the homogeneity of the group and of the environmental milieu, or, on the other hand, to what extent they are related to the facts that the subjects are children of a given age, or children in a specific culture, or that they are simply children. Speculations here would perhaps be promiscuous. There are, however, some data presented by Barker and Wright (1954, Ch. 12) which offer possibilities for very crude comparison. These authors show rankings of interactive categories for the behavior of eight "normal" children, each of whom was observed individually for a day in the small Midwestern town in which the children lived. All interactive behavior was noted and later coded. Four of the children were boys and four were girls; ages range from 1 year, 10 months through 10 years, 9 months, with a median age of 5 years, 10 months. Thus, the children and their environments were very different from those of the present study. The behavioral categories employed in the two studies also

differed, but on an a priori, face-validity basis, it was decided which categories were comparable in the two schemes.¹³ The rank orders of interactive behavior for these two different samples of children in different environments were correlated. Behavior toward adults yielded rhos of .24 and .31 between the Barker and Wright data (1954, p. 429) and the present study's early and late phases, respectively; behavior toward children yielded rhos of .43 and .29. Although none of the rhos is significant with eight sets of ranks, all are positive.

By themselves, positive correlation coefficients can be artifacts of the categories employed; for example, by choosing walking and flying as categories for classification, it can be shown—not wholly illegitimately—that people are all alike. The matter is, however, not so readily dismissed. In the course of the day, the Barker and Wright children had transactions with adults; the same was true for the six boys in the present study. When the total adult behaviors toward the Barker and Wright children (1954, p. 425) were compared in rankings with the total adult behaviors directed toward the six hyperaggressive boys, the rhos were .93 ($p < .01$) for the early phase and .79 ($p < .05$) for the later phase. That is, despite the differences in children in age and psychological status, despite the differences in environments (home, school, and neighborhood versus a psychiatric milieu), despite the differences in adults (parents, neighbors, and teachers as against counselors, nurses, physicians, and teachers), and despite differences in methods of study, patterns of behavior of adults toward children apparently have much in common. The patterns of behavior of the very different samples of children would seem to have less—though still something—in common. In isolation, such findings are difficult to evaluate. They do, however, point to the desirability of investigating commonalities and differences in interpersonal patterns in relation to maturational factors and in relation to cultural and subcultural variations.

¹² The critical value for significance at $p = .01$ (one-tailed test) is a rho of .60.

¹³ Dominance in the Barker and Wright scheme was considered equal to the sum of A and P (Fig. 1), Appeal = K and J, Resistance = B, C, and F, Nurturance = N and O, Aggression = D and E, Submission = I, Compliance = L, and Avoidance = G and H. Affiliative responses, M, were ignored, since affection is treated under a different coding system by Barker and Wright (1954, Ch. 10).

Individual Differences

Another side of the coin is the matter of individual differences. The side one emphasizes, resemblances or differences, depends on what one wishes to talk about. General arguments as to which are greater are specious, a pitfall researchers have not always avoided. For a scheme to be maximally useful, it should be capable both of abstracting general phenomena and of demonstrating differentiations, and, all other things being equal, the finer its discriminations, the greater its potential. An initial and minimal test of the clinical utility of an instrument is its capacity for showing individual differences. Considering the earlier observations by quadrants, the six children differed among themselves in both interactions toward each other (Table 1, $p < .001$) and in interactions toward adults (Table 2, $p < .01$). The "intensity" codings seemed less sensitive to individual factors (Table 3, $p < .01$ in interactions with children; Table 4, $p < .10$ in interactions with adults). Individual differences among the children also appeared in the later phase in interactions toward peers (Table 5, $p < .01$) and in interactions toward adults (Table 6, $p < .05$), but they seemed somewhat attenuated, perhaps by the long period of close communal living. The "intensity" codings yielded results similar to the earlier phase (Table 7, $p < .01$ for interactions with peers; Table 8, $p < .50$ for interactions

TABLE 1

INDIVIDUAL DIFFERENCES IN MODES OF INTERACTION
TOWARD PEERS—EARLY PHASE

(N = 568)

	Hostile-dominant Actions	Hostile-passive Actions	Friendly-passive Actions	Friendly-dominant Actions
Tony	19 (38.97) ^a	24 (11.49)	35 (24.78)	24 (26.76)
Bruce	33 (23.69)	8 (6.99)	10 (15.06)	11 (16.26)
Clif	37 (37.44)	3 (11.04)	26 (23.81)	32 (25.71)
Dave	34 (37.44)	12 (11.04)	17 (23.81)	35 (25.71)
Ed	26 (25.60)	10 (7.55)	18 (16.28)	13 (17.58)
Frank	68 (53.87)	7 (15.89)	32 (34.26)	34 (36.99)

^a Expected values based on marginal distributions are in parentheses. $\chi^2 = 60.13$, $df = 15$, $p < .001$.

TABLE 2

INDIVIDUAL DIFFERENCES IN MODES OF INTERACTION
TOWARD ADULTS—EARLY PHASE *

(N = 542)

	Hostile-dominant Actions	Hostile-passive Actions	Friendly-passive Actions	Friendly-dominant Actions
Tony	24 (24.96) ^a	10 (12.97)	40 (36.29)	15 (14.78)
Bruce	23 (24.96)	25 (12.97)	33 (36.29)	8 (14.78)
Clif	15 (15.70)	4 (8.16)	27 (22.83)	10 (9.30)
Dave	13 (18.51)	6 (9.62)	36 (26.91)	11 (10.96)
Ed	42 (40.10)	24 (20.84)	45 (58.31)	32 (23.75)
Frank	35 (27.76)	10 (14.43)	40 (40.37)	14 (16.44)

^a Expected values based on marginal distributions are in parentheses. $\chi^2 = 34.94$, $df = 15$, $p < .01$.

TABLE 3

INDIVIDUAL DIFFERENCES IN "INTENSITY" OF INTERACTION
TOWARD PEERS—EARLY PHASE

(N = 568)

	Uninvolved Actions	Involved-inappropriate Actions	Involved-inappropriate Actions
Tony	12 (16.34) ^a	86 (76.50)	4 (9.16)
Bruce	14 (9.93)	40 (46.50)	8 (5.57)
Clif	24 (15.70)	72 (73.50)	2 (8.80)
Dave	14 (15.70)	70 (73.50)	14 (8.80)
Ed	7 (10.73)	52 (50.25)	8 (6.02)
Frank	20 (22.59)	106 (105.75)	15 (12.66)

^a Expected values based on marginal distributions are in parentheses. $\chi^2 = 24.71$, $df = 10$, $p < .01$.

with adults), but again perhaps somewhat attenuated.

In all these analyses, individual differences in behaviors toward children seem greater than do individual differences in behaviors toward adults. This trend would indicate that those failures to find significant differences in peer relations between the two phases, as compared to the significant phase differences in relations with adults, were not a function of the lack of sensitivity of the method to peer interactions. On a theoretical level, the trend may point to a more general issue of role relation-



TABLE 4

INDIVIDUAL DIFFERENCES IN "INTENSITY" OF INTERACTION TOWARD ADULTS—EARLY PHASE
($N = 542$)

	Uninvolved Actions	Involved-appropriate Actions	Involved-inappropriate Actions
Tony	14 (14.61) ^a	69 (65.19)	6 (9.20)
Bruce	18 (14.61)	59 (65.19)	12 (9.20)
Clif	5 (9.20)	47 (41.02)	4 (5.79)
Dave	14 (10.84)	50 (48.34)	2 (6.82)
Ed	23 (23.48)	97 (104.74)	23 (14.77)
Frank	15 (16.26)	75 (72.51)	9 (10.23)

^a Expected values based on marginal distributions are in parentheses. $\chi^2 = 16.83$, $df = 10$, $p < .10$.

TABLE 5

INDIVIDUAL DIFFERENCES IN MODES OF INTERACTION TOWARD PEERS—LATER PHASE
($N = 430$)

	Hostile-dominant Actions	Hostile-passive Actions	Friendly-passive Actions	Friendly-dominant Actions
Tony	9 (18.98) ^a	13 (6.84)	20 (17.58)	18 (16.60)
Bruce	25 (16.76)	5 (6.04)	11 (15.53)	12 (14.67)
Clif	24 (23.40)	15 (8.43)	17 (21.68)	18 (20.48)
Dave	20 (26.57)	6 (9.57)	26 (24.61)	32 (23.25)
Ed	28 (24.99)	6 (9.00)	24 (23.15)	21 (21.86)
Frank	30 (25.30)	4 (9.12)	28 (23.44)	18 (22.14)

^a Expected values based on marginal distributions are in parentheses. $\chi^2 = 36.88$, $df = 15$, $p < .01$.

ships. We may speculate that people will tend to act in a more individualized fashion toward those in their own status group than they do toward groups of rather different statuses. That is, not only may there be a cognitive tendency for one group to stereotype another group of very different status, but patterns of behavior in a group may be less variable in relations to other groups than in relations to peers. To the other-status recipient of behavior, the behavioral cohesiveness of a group may then appear to be greater than it actually does to the within-group members.

TABLE 6

INDIVIDUAL DIFFERENCES IN MODES OF INTERACTION TOWARD ADULTS—LATER PHASE
($N = 438$)

	Hostile-dominant Actions	Hostile-passive Actions	Friendly-passive Actions	Friendly-dominant Actions
Tony	9 (10.96) ^a	15 (8.50)	64 (61.08)	10 (17.45)
Bruce	13 (8.61)	8 (6.68)	41 (47.99)	15 (13.71)
Clif	6 (7.83)	8 (6.07)	44 (43.63)	12 (12.47)
Dave	6 (6.82)	2 (5.29)	40 (38.02)	13 (10.86)
Ed	3 (7.50)	3 (5.81)	44 (41.76)	17 (11.93)
Frank	12 (7.27)	2 (5.64)	40 (40.51)	11 (11.58)

^a Expected values based on marginal distributions are in parentheses. $\chi^2 = 27.81$, $df = 15$, $p < .05$.

TABLE 7

INDIVIDUAL DIFFERENCES IN "INTENSITY" OF INTERACTION TOWARD PEERS—LATER PHASE
($N = 430$)

	Uninvolved Actions	Involved-appropriate Actions	Involved-inappropriate Actions
Tony	2 (6.84) ^a	53 (50.51)	5 (2.65)
Bruce	4 (6.04)	42 (44.62)	7 (2.34)
Clif	8 (8.43)	64 (62.30)	2 (3.27)
Dave	11 (9.57)	73 (70.72)	0 (3.71)
Ed	13 (9.00)	65 (66.51)	1 (3.49)
Frank	11 (9.12)	65 (67.35)	4 (3.53)

^a Expected values based on marginal distributions are in parentheses. $\chi^2 = 24.41$, $df = 10$, $p < .01$.

Some Dynamic Aspects of Interpersonal Processes

In all the foregoing discussion, it is as though each child were an independent, self-determined unit. Yet one of the basic facts of interpersonal processes is the interdependence and continuity of behavior. One's actions evoke actions by others, and the actions of others, in turn, stimulate one's own behavior. Leary (1957, Ch. 7) has commented on and given some examples of reciprocity in interpersonal behavior among adults. For cues as to this issue, the behavior that each child received

TABLE 8

INDIVIDUAL DIFFERENCES IN "INTENSITY" OF INTER-ACTION TOWARD ADULTS—LATER PHASE
($N = 438$)

	Uninvolved Actions	Involved- appropriate Actions	Involved- inappropriate Actions
Tony	10 (9.84) ^a	84 (83.90)	4 (4.25)
Bruce	7 (7.74)	65 (65.92)	5 (3.34)
Clif	4 (7.03)	60 (59.93)	6 (3.04)
Dave	7 (6.13)	53 (52.23)	1 (2.65)
Ed	10 (6.73)	55 (57.36)	2 (2.91)
Frank	6 (6.53)	58 (55.65)	1 (2.82)

^a Expected values based on marginal distributions are in parentheses. $\chi^2 = 9.55$, $df = 10$, $p < .50$.

from his peers and from adults was examined. Tables 9 through 12 present the frequencies of interactions and percentages in the four categories of response for behavior each child received from others.¹⁴

Reciprocity in adult behavior toward children. The behavior of the adults contrasted sharply with that of the children. The majority of the adult interactions—a mean over the six children of 58% in the early phase and 72 in the later—were in the friendly-dominant category, composed of friendly, nurturant, supporting, giving, and guiding activities. A further demonstration of the contrast lies in the comparison between behavior "sent" by the children and the behavior they "received." In the early phase, a mean of 51% of the responses which individual children "sent"

¹⁴ These tables were constructed using the interactions of the children when they were *not* in the central focus of observation. That is, in tabulating, the interactions of the central subject were ignored; the interactions of all other persons, children and adults, were tabulated according to whom they were directed toward. This arbitrary rule insures that the same data which entered into the previous tables (1 through 8) do not enter into these tables. At the same time, it should be noted that the entries in any cell do not necessarily represent the equal contribution of all participants to the behavior directed toward a given child. For example, Tony may have received 10 hostile-dominant actions from Bruce and 5 hostile-dominant actions from Frank. Because of such possibilities of confounding, these data are not analyzed by chi square. The discussion that follows is based on visual inspection of the data.

TABLE 9

ACTIONS BY PEERS TOWARD EACH CHILD—EARLY PHASE
($N = 1034$)

	Hostile-dominant Actions		Hostile-passive Actions		Friendly-passive Actions		Friendly-dominant Actions	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Tony	74	46	9	6	33	21	44	27
Bruce	51	37	16	12	38	28	31	23
Clif	65	35	18	10	57	31	44	24
Dave	96	42	29	13	54	24	48	21
Ed	50	40	11	8	27	22	37	30
Frank	78	39	31	15	56	28	37	18

TABLE 10

ACTIONS BY ADULTS TOWARD EACH CHILD—EARLY PHASE
($N = 557$)

	Hostile-dominant Actions		Hostile-passive Actions		Friendly-passive Actions		Friendly-dominant Actions	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Tony	16	18	1	1	8	9	65	72
Bruce	19	23	0	0	13	15	52	62
Clif	24	33	1	1	12	17	35	48
Dave	20	20	2	2	12	12	66	66
Ed	16	16	5	5	28	27	54	52
Frank	26	24	6	6	23	21	53	49

TABLE 11

ACTIONS BY PEERS TOWARD EACH CHILD—LATER PHASE
($N = 920$)

	Hostile-dominant Actions		Hostile-passive Actions		Friendly-passive Actions		Friendly-dominant Actions	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Tony	37	42	6	7	20	23	25	28
Bruce	37	26	7	5	51	37	44	32
Clif	91	41	11	5	58	27	58	27
Dave	36	23	15	9	61	38	47	30
Ed	56	40	19	13	31	22	35	25
Frank	59	34	22	12	54	31	40	23

TABLE 12

ACTIONS BY ADULTS TOWARD EACH CHILD—LATER PHASE
($N = 658$)

	Hostile-dominant Actions		Hostile-passive Actions		Friendly-passive Actions		Friendly-dominant Actions	
	Freq.	%	Freq.	%	Freq.	%	Freq.	%
Tony	19	16	0	0	10	8	91	76
Bruce	18	15	2	2	16	13	83	70
Clif	20	19	0	0	8	7	78	74
Dave	6	8	0	0	13	18	55	74
Ed	8	7	5	4	22	19	82	70
Frank	22	18	1	1	17	14	82	67

toward each other were hostile in orientation (Table 1); they "received" a mean of 51% hostile actions from each other (Table 9). For the later phase, the values were 44% hostile responses "sent" and 43% "received" (Tables 5 and 11). In contrast was the situation with adults. Whereas the children in the early phase "sent" a mean of 42% hostile actions to adults, they "received" a mean of only 25% hostile actions in return (Tables 2 and 10). For the later phase, the values were 19% "sent" and 15 "received" (Tables 6 and 12). Although the data are, of course, not definitive, they hint that processes of interpersonal change, insofar as change occurred, were adult rather than child initiated. It would seem reasonable that if interaction were purely reciprocal—an eye for an eye—one could not expect changes in the direction of the participants' actions, at least after a stable status order had been established. It is probably at the interruptions in patterns of reciprocity that the potential for interpersonal change arises. This is not to say that such interruptions are sufficient requirements for change, although they may be necessary ones.

Nor does the foregoing discussion imply that the adults did not, in some measure, respond reciprocally to the children. Although the adults may have initiated the process of change, they, like the children, showed systematic shifts in behavior between the two phases. The proportions of aggression expressed by adults decreased in relation to each of the six boys.¹⁵ Noted previously was the fact that the major single shift in the 16 categories on the part of the children in their behavior toward adults was the increase in help-requesting, dependent actions (K); it is in-

teresting that the major single category of change in adult responses over the two phases was the increase in giving and help-offering responses (O) from 23 to 34% of the total adult behaviors. The adults also changed in the "intensity" of behavior toward the children. The proportions of appropriate-involved behaviors as compared to uninvolved and inappropriate-involved behaviors increased in relation to each child, the means going from 82 to 90%.

Reciprocity in peer behavior. Changes in actions "received" from peers also, in general, paralleled the previously noted changes in actions "sent" toward peers. The evidence for a general systematic change in the quality of interactions toward peers was noted as slight. Similarly, there appears to be little over-all change in the quality of interactions received from peers (Tables 9 and 11). Dave was the child who changed most in the direction of decreased aggression toward peers. Other children changed similarly in their behavior toward him. A further parallel appears in the case of Clif, who, as mentioned previously, was unusual in that while he decreased in dominant forms of aggression toward peers, he showed significantly more passively hostile actions in the later phase. In the later as compared to the earlier period, Clif received more dominantly hostile but fewer passively hostile actions from peers—the only instance where this occurred. "Intensity" changes in "received" behavior parallel "intensity" changes in "sent" behavior. In all six cases the proportions of appropriate-involved actions increased. But there are deviations from parallelism. For example, there is little evidence that Bruce changed much in his actions toward peers. They, however, seemed to change toward him. He received less aggression and less inappropriate behavior in the later phase.

In general, then, there is a mutuality in interpersonal processes of change. As interpersonal behavior changes, new equilibrium relations tend to form between the person and others. Unfortunately, the study can say all too little about the circumstances under which changes in equilibria occur, or about such problems as lags in mutual rearrangements. It would also seem that realignments may occur on bases which this study missed.

¹⁵ The mutual changes raise the possibility of an interesting counter-argument that the children do not "really" change at all; rather, they simply respond reciprocally to "improvement" in the adults brought about through training and experience. An adequate refutation of this view as a complete explanation of the changes in the children's relations with adults would probably require data extraneous to the present study, although study of the sequences of interactions would provide some cues. Note that such a counter-argument can only arise when there is actual information about the behavior of the "others." The usual assumption in studies of change is of a constant or randomly fluctuating environment. The data put into question the tenability of such an assumption.

One would guess that such factors as interpersonal skills and sensitivities, special cognitive abilities, stability or erraticness of behavior, special friendship and leadership patterns all play a part.

What can be said about the specific nature of reciprocal action? Obviously, nothing definitive, but there are some cues. Let us look, for example, at Tony's relations with the other boys during the early phase. Tony was the least dominantly aggressive and the most passively aggressive of the children. Complementarily, Tony received the greatest proportion of dominant aggression and the lowest proportion of passive aggression from his peers.¹⁶ While Tony initiated the highest proportion of friendly-passive responses to other children, he received the lowest proportion of such responses from peers but the second highest proportion of friendly-dominant behaviors. In contrast to Tony, Frank "sent" a high proportion of dominantly aggressive and a low proportion of passively aggressive behaviors toward the other boys and received from them a high proportion of passive-hostile and a relatively low proportion of dominant-hostile responses. But, as seems true for reciprocity in change, patterns of complementarity in action are not equally clear for all children.

That passive aggression evokes dominant aggression and that dominant aggression evokes passive aggression is not unexpected. What is more interesting is a tendency in the data for passive aggression to evoke friendly-dominant behavior, and for dominant aggression to evoke friendly-passive behavior. The evidence is obviously meagre and the indications are not wholly consistent, but it would seem that hostile oriented actions were not always met with counterhostility. Do "successes" maintain the behaviors? If so, how is the alternative cycle of reciprocal aggression to be counteracted?

SUMMARY

An exploratory study was made of the interpersonal behavior of six hyperaggressive boys in residential treatment. Each child was ob-

¹⁶ Information for these and the following comments may be reconstructed from the tables presented.

served twice in six life settings and his interactions with both peers and adults were noted. The observations were repeated after a year and a half in the treatment program.

Over the year and a half the interpersonal behavior of the children shifted considerably. The major changes were in the relations of the children with adults. Here, there was primarily a decrease in hostile-dominant behavior and an increase in friendly-passive behavior. The appropriateness of behavior increased both in relations with children and with adults. The patterns of change were consistent with treatment aims, and they seemed, at least in part, a function of the treatment program.

Group similarities and individual differences among the children were noted, and patterns of reciprocity in behavior between children and adults and among children were explored. In relations with peers, the children received about the same amount of aggression as they expressed. They received less aggression from adults than they expressed toward them. Changes in patterns of behavior toward others were accompanied, in general, by reciprocal changes in the behaviors of others, both adults and children.

The study demonstrates (a) that systematic observation and coding of the interpersonal behavior of a small group of children in naturalistic settings can yield tenable descriptions, orderly relationships, and some tentative hypotheses about interpersonal processes, (b) that hyperaggressive children can change in residential treatment in a direction consistent with therapeutic aims, and (c) that the mode of observation described here, together with the scheme for coding interpersonal behavior (Freedman et al., 1951; Leary, 1957), has some measure of utility.

REFERENCES

- BALDWIN, A. L. *Behavior and development in childhood*. New York: Dryden, 1955.
- BALES, R. F. *Interaction process analysis: A method for the study of small groups*. Cambridge, Mass.: Addison-Wesley, 1950.
- BARKER, R. G., & WRIGHT, H. F. *Midwest and its children*. Evanston, Ill.: Row, Peterson, 1954.
- BRUNSWIK, E. *Systematic and representative design of psychological experiments*. Berkeley, Calif.: Univer. Calif. Press, 1947.
- CRONBACH, L. J., & MEEHL, P. E. Construct validity

- in psychological tests. *Psychol. Bull.*, 1955, **52**, 281-302.
- DITTMANN, A. T. Problems of reliability in observing and coding social interactions. *J. consult. Psychol.*, 1958, **22**, 430.
- FREEDMAN, M. B., LEARY, T. F., OSSORIO, A. G., & COFFEY, H. S. The interpersonal dimension of personality. *J. Pers.*, 1951, **20**, 143-161.
- HARE, A. P., BORGATTA, E. F., & BALES, R. F. (Eds.) *Small groups—Studies in social interaction*. New York: Knopf, 1955.
- HEYNS, R. W., & LIPPITT, R. Systematic observational techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. Pp. 370-404.
- LEARY, T. *Interpersonal diagnosis of personality*. New York: Ronald Press, 1957.
- REDL, F., & WINEMAN, D. *The aggressive child*. Glencoe, Ill.: Free Press, 1957.

Received August 1, 1957.

HOMOGENEITY OF MEMBER PERSONALITY AND ITS EFFECT ON GROUP PROBLEM-SOLVING¹

L. RICHARD HOFFMAN²

Institute for Social Research, University of Michigan

PSYCHOLOGISTS have only recently begun to show an interest in the relationships between the structure of groups and their problem-solving effectiveness. Rosenberg, Erlick, and Berkowitz (9) have demonstrated an "assembly effect" in the group product, an effect based not on the characteristics of each person but on the relations of the characteristics of each individual to those of the others in the group. Schutz (10) showed that the performances of "compatible" groups were superior to those of "incompatible" groups on two tasks. In these studies, as in most problem-solving experiments, the nature of the tasks used was undefined. Certain questions remain unanswered: Does an "assembly effect" occur on all tasks? Are "compatible" groups superior on every type of problem?

The present experiment examined the interrelationships among different group structures and different task characteristics, as these interrelationships determined the "quality" and "acceptance" of the group product (6). The groups varied in the degree of similarity or homogeneity of their members' personalities. The effects of these variations were examined in terms of the quality and acceptance of solutions to two different problems. One problem involved quality alone; the other involved acceptance and quality.

DEVELOPMENT OF HYPOTHESES

Both Maier (5) and Duncker (1) have emphasized the importance of the "direction" or initial orientation toward a problem in the problem-solving behaviors of individuals.

¹ This article has been adapted from a dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy at the University of Michigan (3). The author wishes to express appreciation to Norman R. F. Maier for his continuing interest and stimulating guidance in the research and in the preparation of this report. Thanks are also offered to Arnold Tannenbaum and Annette Wigod for constructive criticisms of this article.

² Now with Engineering Research Institute and the Department of Psychology, University of Michigan.

Maier presented evidence that a person who has many "directions" available, i.e., is capable of many restructurings of his perceptual field, is more likely to be a successful solver than the person who is inflexible and adheres to a single direction.

On the assumption that "direction" operates in the same way in groups as in individuals, the variety of possible directions available from the members of Nonhomogeneous groups, deriving from their varied perceptual and cognitive structures, should yield higher quality solutions in these groups than in Homogeneous groups. The higher quality solutions should result from the many different ideas and from the emphasis placed on different aspects of a problem by the members of Nonhomogeneous groups. This conclusion should be true only for those tasks requiring multiple perceptions and cognitive reorganizations for their solution. Hypothesis I states therefore: On a problem involving quality alone, Nonhomogeneous groups produce higher quality solutions than do Homogeneous groups.

The nature of Nonhomogeneous groups implies differences in affective structures—in the innerpersonal regions—as well as in cognitive structures among the group members. When a problem has no objectively good solution—i.e., when the quality of the solution can be judged only in terms of the members' personal values and standards—then differences in affective structures in Nonhomogeneous groups should again produce conflict. The result of such conflict is difficult to predict. If the values involved are not central to the members of the group there is a high probability of affective restructuring and easy resolution of the conflict. The group is likely to generate a unique and interesting solution. If the issues are more central, affective restructuring is less likely to occur, and the Nonhomogeneous group is more likely to fail to agree on a solution. The status quo situation will continue. Problems requiring affective restructuring are considered to be *acceptance* problems.

To the extent that a problem involves both quality and acceptance, successful solutions require changes in both cognitive and affective structures of the group members. Hypothesis II therefore states: On a problem involving quality and acceptance, Nonhomogeneous groups either fail to agree on a solution and resist change, or produce inventive solutions. Homogeneous groups produce few inventive solutions, but either fail to agree or accept an alternative solution offered to them.

Hypothesis III, concerning satisfaction with the solutions, also derives from the preceding considerations. By assumption, a solution is acceptable to a person to the extent that it satisfies his needs in the situation; in a problem-solving situation a good solution satisfies certain of his needs. If he then perceives that the best suggestions made by the members are incorporated in the group's solution, he should be satisfied with the solution. In the Homogeneous groups, approval should be almost unanimous. The solutions, although incorporating only a limited number of ideas, represent the opinions of all the group members, since these should tend to be similar. In Nonhomogeneous groups, however, unanimous satisfaction should be found only in those groups with high quality solutions, where the multitude of ideas produced have been used in the solution. On this basis, Hypothesis III states: For Homogeneous groups, there is no relationship between the degree of satisfaction with the solutions and their quality for either type of problem. For Nonhomogeneous groups, there is a positive relation between the degree of satisfaction with and the quality of the solutions.

METHOD

The experimental procedure was conducted twice, once in the spring semester and again in the fall semester of 1955. The subjects (Ss) were sophomore, junior, and senior students in the undergraduate course, psychology of human relations. The 175 students were assigned to seven laboratory sections of about 25 students each. In laboratory sessions, the students participated in a case discussion or role-playing situation each week to learn skills in interviewing and in group leadership.

Personality Measure

The Guilford-Zimmerman Temperament Survey (GZTS) (2) was administered as the personality measure at the beginning of each semester to all stu-

dents. Although the GZTS was developed to measure personality traits and not tension systems in people it was selected as the personality measure for its reliability and for the relative independence of its dimensions. It was assumed that the ten traits it measures provide a sample of the personality characteristics such that Ss with similar profiles are likely to be more similar in personality than those with dissimilar profiles. Kendall's tau was used to determine the correlation between the ten-score profiles of every pair of students in each laboratory section. Approximately 2,000 correlations were computed each semester for the seven laboratory sections. With the aid of the Michigan Digital Automatic Computer (MIDAC), these computations were made in sufficient time to establish the groups by the following week's meeting.³

The reliability of the profile correlations was computed by a split-half technique on a sample of 50 Ss. Corrected Spearman rank-order correlation coefficients between odd and even halves of the GZTS had a median of .77. Thirty-seven of the 50 coefficients equaled or exceeded .64, the value needed for significance at the .05 level of confidence.

Experimental Groups

Three types of four-person groups were assembled. Type 1 groups (Homogeneous) were composed of persons with high positive taus; Type 2 groups were composed of persons with both high negative and high positive profile correlations; and Type 3 groups consisted of persons whose taus were approximately .00. The number of groups of each type initially established were, respectively, 20, 15, and 22 for a total of 57 groups. A comparison of the algebraic and absolute sums of tau for the three types of groups provided statistical evidence that there were distinct qualitative differences in the personality mixtures of each group type. The Homogeneous groups, moreover, represented such different dominant personality patterns that a subsequent attempt to analyze possible differences among groups of this type had to be abandoned for lack of a sufficient number of groups of any one personality type. Although students were assigned to these different types of groups purely on the basis of their personality characteristics, comparisons of the various types of groups also indicated no differences in the sex composition (the mixture of males and females) and in the mean final course grade.

Dependent Variables

The solutions to two problems and measures of satisfaction with these solutions served as the dependent measures. The first of these problems—one

³ The author thanks his wife, Roslyn B. Hoffman, for programming the computation of tau for the MIDAC and John W. Carr, III and Cecil Craig representing the Rackham School of Graduate Studies for allowing me to use the MIDAC and for providing funds to pay for the computer time required. Without this assistance the magnitude of the computations involved would have made this study impossible to do within the limited time available.

involving quality alone—was the Mined Road Problem (MRP).⁴ It was administered in both semesters in the next to the last meeting of the semester.

On this problem, the group is told that it is a five-man guerilla team which has just blown up an enemy bridge and is attempting to return to its own lines. In order to meet a truck that is to carry the group back to the base, the men must cross a road that is known to be heavily mined. Scattered around the area are some scrap materials, e.g., ropes, lumber, etc. that could be used to cross the road. The problem is to determine the best method for crossing the road safely, quickly, and without traces of crossing.

Solutions to the problem were assigned a numerical score by a content analysis scheme developed by Lorge and his associates (4). A solution accumulated points according to the relative feasibility of the method used to cross the road, the relative safety, the concealment of clues, and the time taken to solve the problem. The reliability of the scoring system was computed by a repeat scoring of a sample of 19 group solutions two weeks after the initial scoring. The correlation between these two sets of scores was .84, high enough for confidence in the results obtained from the use of the scoring system.

The second problem was the Change of Work Procedure (CWP) problem developed by Maier for use in training supervisors in human relations practices (6, p. 54). The problem was considered primarily an affective problem with a high "acceptance" component (6). That is, although solutions to the problem can be classified into different types, there is no objectively "best" solution; the best solution is that most acceptable to the group, the one they would be willing to carry out. Because certain solutions can be considered to be inventive, however, the problem also has a quality component.

The problem is a role-playing situation in which three workers, who perform three different jobs in hourly rotation with each other, report to a supervisor, the fourth man in the group. The supervisor, at the suggestion of a time-study man, requests the group to work fixed positions instead of rotating their jobs, to achieve more efficient production. Solutions to the problem resolve the conflict between attempts to capitalize on individual ability for higher production (supervisor's suggestion) and the freedom from monotony resulting from rotation. Three types of solutions are usually obtained: Old solutions: the group refuses to change or continues rotation with a different time interval between changes; New solutions: complete acceptance of the supervisor's suggestion or acceptance with some minor modifications like rest pauses or music; and Compromise or Inventive solutions: work procedures, not included in the roles themselves, that attempt to gain the advantages of both individual ability and freedom from monotony. Assignment of the solutions obtained in this study to these categories offered no difficulty. A reclassification of the solutions

after a two-week interval showed almost perfect agreement with the initial classification. The scorings of both MRP and CWP were done without knowledge of what type of group had submitted the solution.

For both problems, measures of the satisfaction with the solution reached by the group were collected. For the MRP, the question took the form, "Were you satisfied with: (a) the entire solution, . . . (d) practically none of the solution?" For the CWP, students were asked, "How satisfied are you with the solution reached by the group? (a) Very satisfied, . . . (e) Very dissatisfied."

The problem-solving activity, the collection of solutions, and the expressions of satisfaction with the solutions were part of the usual classroom procedure. The experimental procedure introduced written solutions and written expressions of satisfaction instead of the usual oral collection of these data.

RESULTS

The results of the problem-solving sessions will be discussed first in terms of the quality and the kinds of solutions obtained in the two groups, and second in terms of the acceptance of the solutions by the groups.

Quality and Kind of Solutions

When the MRP was administered, only four Type 2 groups were wholly intact. These four were combined with the 13 intact Type 3 groups to form 17 Nonhomogeneous groups for the analysis. To test Hypothesis I, the scores for these groups have been compared with the scores of the 13 intact Type I (Homogeneous) groups in Table 1.⁵ The higher the score the better is the solution. The significant 18.6 point superiority of the mean of the Nonhomogeneous groups over that of the Homogeneous groups supports Hypothesis I. The superiority of the Nonhomogeneous groups is even more apparent when the two distribu-

⁵ The data for both problems are based on the products of only those groups which were wholly intact (i.e., all four members were present) at the time the problem was administered to the class. The use of only intact groups for the analysis attempted to avoid bias due to differences in group size, and to any non-random forces that might be operating to cause a person to be missing (e.g., rejection by or of the group) and to make the intact parts of these groups different from the wholly intact groups. Sixty-five per cent and 60% of the Homogeneous groups were intact for the MRP and the CWP respectively. Forty-six per cent of the combined Nonhomogeneous groups were present for both problems. This percentage difference is not statistically significant. A significantly smaller percentage (27%) of the Type 2 than Type 1 groups was intact for the MRP.

⁴ I wish to thank Irving Lorge of Columbia University for supplying me with copies of the Mined Road Problem and the scoring key and for discussing the interpretation of the scoring procedure with me.

TABLE 1
COMPARISON OF SCORES ON MINED ROAD
PROBLEM

Group Type	Tau	Mean*	SD	Number of Groups
Homogeneous	+	44.5	28.63	13
Nonhomogeneous	0, -	63.1	28.53	17

* One-tailed *t* test that the mean of Nonhomogeneous groups is greater than the mean of the Homogeneous groups is significant at the .05 level.

TABLE 2
NUMBER OF GROUPS PRODUCING EACH
TYPE OF SOLUTION

Group Type	Tau	N	Type of Solution		
			Old	New	Inven- tive
Homogeneous	+	12	4	6	2
Nonhomogeneous	0, -	17	2	8	7

tions are compared. When the results of all the groups were combined, only 3 (23%) of the 13 Homogeneous groups exceeded the median score for the total distribution, whereas 12 (71%) of the seventeen Nonhomogeneous groups surpassed that score.

Although the groups were assembled purely on the basis of the degree of homogeneity of the members' personalities, the question may be raised as to whether the obtained differences can be safely attributed to the personality differences alone. Three other variables which could have accounted for the differences were examined: (a) intelligence and knowledge of the subject matter of the course, as measured by the final course grades, (b) the sex composition of the group—the number of women in the group, and (c) the sociometric attractiveness of the group. No significant differences were found to distinguish the groups, thus reinforcing the causal link between the personality composition of the groups and their problem-solving qualities.

The results for the CWP were less clear-cut. The distributions of solutions for the Homogeneous and Nonhomogeneous groups are presented in Table 2. The chi-square value for the contingency table comparing the Homogeneous and Nonhomogeneous groups is 2.88, which is not significant at the .05 level of confidence. Hypothesis II was therefore not confirmed. Both Homogeneous and Nonhomogeneous

groups tended to accept the supervisor's suggested work method, the New solution.

Considering Inventive solutions alone, however, only 2 (16%) of the 12 Homogeneous groups as against 7 (41%) of the 17 Nonhomogeneous groups produced such solutions. Although the percentage difference is not statistically significant, it suggests that the resolution of conflict generated in certain Nonhomogeneous groups resulted in ideational creativity.

Acceptance of Solutions

The data reflect a high degree of satisfaction with the solutions to the MRP in both types of groups, despite the fact that the solutions to the problem produced by the Homogeneous groups were so inferior by the objective criterion. In 21 of the 30 groups, all four people agreed with the final group solution. The tendency toward group acceptance of the solution was especially marked in the Homogeneous groups, where the mean number satisfied was 3.8 out of a possible 4.0, and where only 2 of the 13 groups were not completely in accord on the final solution. In the Nonhomogeneous groups, the mean was 3.3 and 7 of the 17 groups were not unanimously satisfied.

The data support the first part of Hypothesis III, which predicted no relationship between the quality of the solution and the degree of satisfaction with the solution in the Homogeneous groups. Where there is no variance there can be no relation. Almost every Homogeneous group was satisfied with the eminently poor solutions they achieved.

However, in the Nonhomogeneous groups satisfaction was not related to the quality of the solution either. When the distribution of scores on the MRP was dichotomized at the median, and the Nonhomogeneous groups whose members were completely satisfied were compared with those where satisfaction was less than unanimous, no relationship was found between the quality of and satisfaction with the solution. The expectation that members of Nonhomogeneous groups would be less satisfied with a poor solution than with a good one was not confirmed by the data.

Hypothesis III was not confirmed by the data for the CWP either. The relationships within each group type, although in the positive direction, are not statistically different

TABLE 3
RELATION BETWEEN TYPE OF SOLUTION
AND SATISFACTION WITH THE
SOLUTION
(Change of work procedure)

Group Type	Type of Solution	
	Old or New	Inventive
Homogeneous		
Number satisfied		
Four	3	2
Less than four	7	0
		$p = .15$
Nonhomogeneous		
Number satisfied		
Four	6	6
Less than four	4	1
		$p = .24$
Total		
Number satisfied		
Four	9	8
Less than four	11	1
		$p = .03$

from each other nor are they significantly different from zero. For both types of groups, where the group generated the Inventive solution, the members of the group were unanimously satisfied with the solution (Table 3). In the one group which produced an Inventive solution but where satisfaction was less than unanimous, three of the four members were satisfied. The relationship for the combined groups is significant at the .05 level. If a group succeeded, on this problem of quality and acceptance, in generating a solution which encompassed both problem aspects of the situation—i.e., obtaining greater productivity without monotony and boredom—the group members accepted the solution as satisfactory.

DISCUSSION

The results of this study support the value of studying the effects of the structure of groups without regard to the particular characteristics of the individuals in the groups. An important determinant of the group's effectiveness is the *interaction* of these individual characteristics. The degree of homogeneity of personality of the members of the groups used in this study was seen to have a direct bearing on the effectiveness of the groups in producing solutions to problems.

As defined for the present study, homogeneity is a relatively pure concept, reflecting the degree of similarity of the personality types comprising the group membership,

rather than the particular personality types of the individuals in the groups. The members of different Homogeneous groups were not interchangeable, and the personality types that were dominant in the Homogeneous groups varied from group to group. The members of the Nonhomogeneous groups obviously varied considerably in the types of personalities they represented. The relative "pureness" of the concept of homogeneity makes all the more impressive the finding that the Nonhomogeneous groups did significantly better on the MRP than did the Homogeneous groups, and that there was a tendency for the Nonhomogeneous groups to produce more Inventive solutions on the CWP.

The results imply that a multiplicity of perceptions of a problem are productive of creative solutions. In the present trend towards group research in scientific organizations and toward group decision-making in various administrative circles, the abilities of each individual may be less important than the peculiar composition of backgrounds and experiences represented by the various members of the team. Pelz, in a study of a scientific research organization (7), found that frequent contact between a scientist and other members of his research group was related to the scientist's own productivity only in those situations where the other members of the group were dissimilar from him in their work motivations and previous work experience. Pelz's findings suggest that the results reported in the present study are probably generalizable well beyond the limited population of college students who supplied the data.

The findings with respect to satisfaction of the group members with the solution to problems also bear comment at this point. On the MRP there was a high degree of satisfaction with the solutions obtained in almost every group regardless of type. Even in those Nonhomogeneous groups in which unanimous satisfaction was not achieved, only one or two members in each group were not completely satisfied with the solution. On the CWP, however, the only groups which showed a similar unanimous satisfaction with the solution were those which produced Inventive solutions. These two sets of results suggest the possible fruitfulness of a distinction between *decision-making*, in which the group members choose among one of several alternatives, and *prob-*

lem-solving, in which the members work through a problem and create a solution. The majority of solutions to the CWP were Old or New solutions, the results of choosing between alternative solutions already given in the problem situation. The creation of Inventive solutions to the CWP and of all solutions to the MRP required the groups to combine elements and develop new procedures. For the college student group in this study, the creative process was a more satisfying experience than the decision-making one.

A final important consideration emphasized by the results of this study is the lack of adequate classification of problem-solving tasks in the psychological literature. An attempt was made to classify the two problems in this study on an a priori basis according to the cognitive or quality aspects versus the affective or acceptance aspects of the problem. The MRP was considered to be purely a quality problem requiring the group members to reorganize their cognitive fields in a way which would provide solutions to the problem. The CWP was considered to be a problem involving both quality and acceptance. The hypotheses concerning differences between the performances of Homogeneous and Nonhomogeneous groups rested on the assumed validity of these classifications. The lack of support for the hypothesis concerning solutions to the CWP (Hypothesis II) may have been a refutation of the theory or possibly only of the validity of the problem description.

Ray (8) recently performed a service for psychologists by bringing together the variety of problems frequently used in research on the problem-solving process. The lack of any apparently common elements in these tasks points up the problem of predicting group performance knowing only the nature of the group. The interrelationships among group and task characteristics are fundamental, and the study of group problem-solving cannot afford to ignore them.

SUMMARY

The study investigated the relationships between the characteristics of groups and of problem-solving tasks, and the effects of these relationships on the groups' problem-solving performances. The performances of groups varying in the homogeneity of their members'

personalities were compared on a problem involving quality alone, and on another involving acceptance and quality.

On the basis of correlations between pairs of individual score profiles on the Guilford-Zimmerman Temperament Survey, groups of four students each were established at the beginning of two semesters in a course in human relations. Homogeneous groups, in which the members had high positive profile correlations, and Nonhomogeneous groups, in which the profile correlations were zero or negative, were maintained throughout each of the semesters.

The Nonhomogeneous groups produced significantly superior solutions to the quality problem, and showed a tendency to produce more inventive solutions to the problem involving quality and acceptance. No significant differences were found in the degree of satisfaction with solutions to either of the problems. On the problem of quality and acceptance, a significant correlation was found in both types of groups between the members' satisfaction with the solution to the problem and the quality of the solution.

REFERENCES

1. DUNCKER, K. On problem-solving. *Psychol. Monogr.*, 1945, 58, No. 5 (Whole No. 270).
2. GUILFORD, J. P., & ZIMMERMAN, W. S. *The Guilford-Zimmerman Temperament Survey*. Beverly Hills: Sheridan Supply, 1949.
3. HOFFMAN, L. R. Homogeneity of member personality and its effect on group problem-solving. Unpublished doctoral dissertation, Univ. of Michigan, 1957.
4. LORGE, I., TUCKMAN, J., AIKMAN, L., SPIEGEL, J., & MOSS, G. Solutions by teams and by individuals to a field problem at different levels of reality. *J. educ. Psychol.*, 1955, 46, 17-24.
5. MAIER, N. R. F. Reasoning in humans. I. On direction. *J. comp. Psychol.*, 1930, 10, 115-144.
6. MAIER, N. R. F. *Principles of human relations*. New York: Wiley, 1952.
7. PELZ, D. C. Some social factors related to performance in a research organization. *Administrative Sci. Quart.*, 1956, 1, 310-325.
8. RAY, W. S. Complex tasks for use in human problem-solving research. *Psychol. Bull.*, 1955, 52, 134-149.
9. ROSENBERG, S., ERICK, D. E., & BERKOWITZ, L. Some effects of varying combinations of group members on group performance measures and leadership behaviors. *J. abnorm. soc. Psychol.*, 1955, 51, 194-203.
10. SCHUTZ, W. C. What makes groups productive. *Hum. Relat.*, 1955, 8, 429-466.

Received August 26, 1957.

THE EFFECTS OF MILD FRUSTRATION ON THE EXPRESSION OF PREJUDICED ATTITUDES^{1, 2, 3}

EMORY L. COWEN

University of Rochester

JUDAH LANDES

Stanford University

AND DONALD E. SCHAEFF

U. S. Marine Corps

THE concept that unacceptable aggressive or hostile impulses may be "displaced" to targets more suitable than the original one has been with us in psychology at least since the writings of Sigmund Freud (8). However, it is primarily as a result of the explicit formulation of frustration-aggression theory (7) that concerted experimental test of this proposition has been attempted in diverse areas. One specific formulation derived from these conceptualizations is that increasing personal frustration may have, as one consequence, an increase in expression of prejudice. Such a theoretical notion has been referred to as a "scapegoat" theory of prejudice (23). A more detailed consideration of possible relations between frustration of personal needs and prejudice has been presented by Krech and Crutchfield (13) under the heading of "a motivational analysis of prejudice."

Criticism has been directed to a scapegoat theory of prejudice both on theoretical and empirical grounds. In the former instance, the argument has been advanced that a scapegoat theory is an insufficient basis for explaining a sizeable number of instances of prejudice (23). As stated, there can be little question as to the justifiability of this argument. On the other hand, a scapegoat theory of prejudice may quite appropriately be viewed as no more than one of a series of explanatory principles required for complete understanding of the phenomena of prejudice. Gordon Allport, in his scholarly treatment of

this problem (2), takes exactly such a position. Allport reviews six major classes of theoretical explanations of prejudice and points out that each seems to constitute a constructive vehicle for augmentation of our understanding of the phenomenon. Allport states "...as a rule most 'theories' are advanced by their authors to call attention to some one important causal factor, without implying that no other causal factors are operating" (2, p. 207). It may therefore be important to re-emphasize that when we are dealing with complex social processes such as, for example, prejudice, delinquency, industrial conflict, and international tensions, multiple determinants are likely to be involved. The identification of a single determinant does not in any way positively demonstrate that this is a sole determinant; nor does it necessarily preclude the operation of differing determinants toward the same end result.

An examination of some empirical data bearing on a scapegoat type theory indicates fairly conclusively that such an explanation should indeed be considered partial. For example, Morse and Allport (17), in a comprehensive investigation of seven hypotheses about the causes of anti-Semitism, found that only the factor of "national involvement" co-varied uniquely with anti-Semitism. "Circumstance frustration," the factor most directly derivable from a scapegoat theory, related only modestly to discriminatory treatment of Jews, leading the authors to conclude that scapegoat theories may not be taken as "general explanations of anti-Semitism."

Lindzey (15), in partial support of a scapegoat explanation, reported that both high and low prejudice Ss increased significantly in displaced aggression following frustration. On the other hand, since the high prejudice

¹ Portions of the present paper were presented to Division 8 at the annual meetings of the American Psychological Association in New York, September, 1957.

² The authors wish to express their appreciation to Russel F. Green for his contributions to the final method of analysis.

³ The study was carried out while both junior authors were at the University of Rochester.

Ss (contrary to deductions) failed to display more aggression than the lows, the author rejects scapegoating as a comprehensive explanatory principle.

Studies offering less qualified support for the existence of the scapegoating phenomenon are also reported in the literature. Thus Allport and Kramer (3), in their classic investigation of the "roots of prejudice," observe that among their Ss (Harvard, Dartmouth, and Radcliffe undergraduates) Catholic and Jewish Ss who saw themselves as more victimized also tended to be more prejudiced toward other minority groups. These findings, interpreted within a frustration aggression framework, were subsequently replicated by Rosenblith (20) with South Dakota undergraduates. Gough (9) found that high anti-Semite Ss are "less able to overlook and ignore minor irritations and frustrations." Mussen (18) reports that high prejudice children had stronger aggressive and dominant needs than did low prejudice Ss, and that they also showed an increase in prejudiced feelings toward Negroes, in contrast to lows who showed a decrease, following four weeks in an interracial summer camp. Finally, Bettelheim and Janowitz (4) have demonstrated significant contingencies between downward social mobility of veterans and intensity of anti-Semitic and anti-Negro attitudes.

With regard to the empirical data thus far reviewed, we may tentatively conclude that a scapegoat concept provides a basis for understanding some instances of prejudice, but is insufficient as a general explanatory principle. Two additional investigations (5, 15), each of which constitutes a direct test of the scapegoat proposition remain to be considered. Miller and Bugelski (16), working in the context of a CCC camp, were able to show a significant drop in positive attitudes and some trend toward increasing negative attitudes toward Mexicans and Japanese following a rather realistic experimental induction of frustration. On the basis of these findings the authors concluded that frustration increased aggression, which was in turn displaced in the form of deterioration of attitude to minority group members.

Recently, however, on the basis of some experimental work by Congdon (5), as re-

ported by Stagner and Congdon (22), some question has been raised with respect to the generality, if not substance, of the Miller and Bugelski findings (15). Congdon assessed attitudes toward various in-groups and out-groups using a series of modified Osgood semantic-differential-type scales. Following this, experimental Ss in two groups were either mildly or strongly frustrated by failure on two of four or four of four subtests of the Grace-Arthur, respectively. A control group received no frustration. Subsequent readministration of the attitude scales indicated no differences in attitude change scores among the three groups. On the basis of these data, Congdon challenges the defensibility of a scapegoat theory of prejudice. He goes on to speculate that the failure to support the Miller and Bugelski findings may reflect some combination of: (a) having used a less arbitrary type of frustration (e.g., see Pastore (19)), (b) having provided outlets for self-punitive behavior which were presumed not to have been present in the Miller and Bugelski experiment, and (c) the higher intellectual level of his subjects.

If on theoretical grounds one espouses, as we have, the view that a scapegoat theory of prejudice may be most useful as one of a series of complementary explanatory principles underlying the complex social phenomenon of prejudice, the Congdon findings raise the question as to whether such a view is useful *even* as a single particularist explanation for understanding some manifestations of prejudice. It is to this latter specific issue that the present research is addressed. In essence, we have attempted to re-examine the proposition that frustration will lead to an increased verbal expression of prejudice, preserving in our design the features of nonarbitrariness, opportunity for expressive and self-punitive behavior, and high intelligence level of Ss, to which Congdon has attributed his negative findings.

METHOD

Instruments

Two comparable subscales, each presumably measuring authoritarian attitudes and minority group prejudice, were drawn from a larger pool of items utilized in the California studies (1). These included the 30-item F scale (combined Forms 40 and 45), the 12-

item Anti-Negro (AN) scale, and eight items each from the Anti-Minority (AM)⁴ and Patriotism (P) scales of the larger Ethnocentrism (E) scale. Items were assigned so that the subtests would be equated with respect to item discrimination quotients as reported in *The Authoritarian Personality* (1). The final forms (X and Q) each contained 29 items as follows: (a) F scale—15 items, (b) AN scale—6 items, (c) AM scale—4 items, (d) P scale—4 items.

Subjects and Procedure

¹Subjects (Ss) were 32 male and 32 female introductory psychology volunteers, all of whom were tested individually. The actual experiment consisted of three phases: the attitude pretest, frustration, and the attitude posttest. In the pretest phase, Ss were given either Form X or Form Q of the attitude scale, with order of presentation counterbalanced for both sexes. Immediately following completion of this first attitude scale, Ss were informed that we wished to collect some additional and separate data bearing on the problem-solving habits of college students. It was in this context that frustration was introduced.

In order to induce frustration, all Ss were given two puzzles which, though appearing soluble, were functionally nonsoluble in the time allotted. The actual puzzles used were the nine-dot problem reported elsewhere by Cowen (6), and one of the Katona match stick problems (12). Fictitious time norms were given, so as to increase the likelihood that frustration would occur. The actual time allotted by E fell far short of what would be needed by most people to solve the problems.⁵ The attitude of the E during administration of the frustrating puzzles might best be described as aloof, non-supporting, and disbelieving of Ss inability to achieve a correct solution.⁶ Upon the Ss failure to solve the second puzzle E stated simply, "I'm afraid our time is up for this problem too. We will have to complete the second part of the attitude scale now." The S then completed the alternate form of the combined attitude scale.

RESULTS

Table 1 summarizes mean "pre" and "post" frustration test scores for each of the attitude test subscales. The data are presented separately by sex and for the two orders of administration.

⁴Several items which were considered inappropriate either because of temporal or geographic factors (e.g., "zoot-suiters" and "Filipinos") were either deleted or slightly modified in wording.

⁵Five Ss who got correct solutions to one of the puzzles, were dropped from the study and replaced by new Ss.

⁶It was the impression of the Es that failure to solve the puzzle did constitute a frustrating experience for the great majority of Ss in terms of such behavioral manifestations as increased fidgetiness or blushing, and verbal comments of discomfort, self-depreciation and/or hostility to the Es.

TABLE 1
MEAN "PRE" AND "POST" SCORES
FOR ALL SUBSCALES

Scale	Sex	Order			
		X pre	Q post	Q pre	X post
AN	M	11.2	15.6	10.4	10.6
	F	9.9	11.3	11.3	10.8
F	M	49.7	44.9	44.0	55.2
	F	47.4	42.6	40.9	46.3
AM	M	8.6	9.4	10.3	9.4
	F	7.8	8.8	7.9	8.1
P	M	13.6	11.7	13.4	13.5
	F	12.4	11.4	12.1	12.9

TABLE 2
ANALYSIS OF VARIANCE FOR ANTI-NEGRO SCALE

Source	df	Sum of Squares	Mean Square	F	P
Between Ss	63	1942.5			
B	1	39.1	39.1	1.33	n.s.
A ^a × C	1	48.5	48.5	1.64	n.s.
A × B × C	1	86.4	86.4	2.93	n.s.
Error (b)	60	1768.5	29.5		
Within Ss	64	345.5			
A	1	73.3	73.3	25.28	.001
C	1	58.0	58.0	20.00	.001
A × B	1	9.8	9.8	3.38	n.s.
B × C	1	30.2	30.2	10.41	.01
Error (w)	60	174.2	2.9		
Total	127	2228.0			

Note.—A = Test form; B = Sex; and C = Pre-Post frustration.

For each of the constituent subscales, a three-way analysis of variance was carried out, involving the main effects of test form (A), sex (B), and pre-post frustration (C). Since both test form and pre-post frustration (the effect in which our major interest centered) are "within subjects" effects, a Lindquist Type IV design (14) was employed as the basic model for the analysis. In general, the only subscale that presented consistently positive findings was the AN scale.

AN Scale

Table 2 presents the results of a three-way analysis of variance for the Anti-Negro scale. The most salient findings in this table are the significant *F* ratios involving pre-post frustration (C). Although the main effect here is

highly significant, suggesting, in support of our hypothesis, the presence of stronger anti-Negro feelings following frustration, this finding may be pinpointed somewhat more specifically by noting other significant main effects and interactions, together with the means presented in Table 1. Thus it appears that there may be differences in the two supposedly equated subtests, with higher AN scores being given on Form Q. Perhaps more germane is the significant $B \times C$ interaction, indicating that male Ss express significantly stronger anti-Negro attitudes after frustration than do female Ss.

Incidental Findings

For the P scale, no significant main effects or interactions are observed. The pattern of findings for the AM and F scales is highly similar. In each case there are significant differences in the two test forms, a finding entirely tangential to our present focus, and on both scales there is either a significant (AM) or near significant (F) main effect of sex. Male Ss tend to score consistently higher (more negative attitudes) on both of these scales. This difference, however, is a general one, which does not vary systematically for "pre" vs. "post" frustration.

In order to test the generality of responses to the various subtests, a series of 12 Pearson product-moment correlations, in which AN scores were related to each of the other three scale scores (by sex, for both the "pre" and "post" tests), were computed. The correlations ranged in magnitude from .27 to .58 and averaged .40. These correlations are substantially lower than the ones reported by the California group (1).

DISCUSSION

The most notable finding in the present experiment is the significant increase in anti-Negro feelings following experimental induction of frustration. Such a datum offers additional support for the existence of the scapegoat phenomenon and is quite consistent with earlier findings of Miller and Bugelski (15). Of incidental interest is the consistent trend observed on three of the subscales for male Ss to show greater prejudice (as well as

greater increase in prejudice following frustration on the AN scale) than do females. This finding too is in line with empirical evidence (3, 20) and theoretical expectations (21) discussed elsewhere.

There remains a sharp contrast between our basic findings in support of a scapegoat theory and those of Congdon (5) which fail to support this view. The latter has proposed that his failure to obtain significant postfrustration effects may reflect some combination of less severe and less arbitrary frustration, provision of opportunities for self-punitive behavior, and the higher intellectual level of his Ss. In the present study, frustration was neither severe nor arbitrary or, at least insofar as can be judged, no more so than Congdon's. Opportunities for self-punitive behaviors should have been roughly comparable, as was the intellectual level of the Ss. These factors notwithstanding, it was possible to demonstrate the operation of the scapegoat effect in the present study.

The source of the discrepant findings in these two ostensibly comparable studies cannot be identified with any confidence, but several procedural variations may be noted that might have obscured the scapegoat effect in the Congdon experiment. The attitude dimensions used for the ratings tended to have quite highly crystallized social desirability values (e.g., *kind—cruel, strong—weak*, etc.). Both pre- and postscores may thus have been pushed toward the socially desirable response, obscuring differences. Then, too, the use of identical items in both pre- and posttest (in contrast to the alternate forms of the present study) may have operated to help sophisticated Ss sense the purpose of the experiment. Finally, Congdon used speed instructions, which may possibly have impaired the reliability of the attitude scales.

Possibly the most interesting issue raised by our findings is the fact that significantly higher anti-Negro attitudes are present following frustration, in the absence of a parallel increase in F, AM, and P scale scores. In the basic development of these scales (1), substantially high intercorrelations have been reported. Whether such correlations reflect a true clustering of these classes of attitudes, or a per-

vasiveness of response tendency behavior, the fact remains that there was reason to anticipate that they would have "behaved" similarly in our study. However, they did not. Our own pre- and postscale intercorrelations run substantially lower than those originally reported, suggesting that our Ss responded with a degree of independence to the subscales. The positive results on the AN scale can be seen most defensibly as an illustration of "targeting" a specific minority group (2, 10, 23). That the Negro is the targeted group in the present study may be a manifestation of a tendency noted earlier by Horowitz (11) for this group to constitute a preferred target in this geographic locale. In another vein, Bettelheim and Janowitz (3) observe that thresholds for anti-Negro prejudice may be lower than those for other minority groups to the point where negative attitudes may break through despite the presence of "relatively adequate controls." In agreement with such an interpretation is our observation that anti-Negro feelings seem to constitute a preferred prejudice in informal conversations of undergraduates at this institution.

In general overview, the present findings are viewed as confirming the hypothesis that frustration augments the expression of prejudice. The major limitation placed upon this conclusion is that the consequent increase in prejudice may be specific rather than generalized. Our findings in no way limit the role or importance of other types of antecedents of prejudice. Undoubtedly the relationships between many such antecedents and the same final product will have to be identified if we are ultimately to have an adequate, comprehensive theory of prejudice. For the present, however, the usefulness of a scapegoat theory, at least as one of a series of complementary explanatory principles, appears defensible.

SUMMARY

The present study was designed to test the proposition that frustration may increase the expression of prejudice. Sixty-four Ss were given a series of attitude scales, following which all were exposed to a relatively mild, experimentally induced frustration. Immediately

thereafter, alternate forms of the attitude scale were administered.

Significant increases were found in the expression of anti-Negro prejudice following frustration, this effect being more pronounced in male Ss. Since comparable postfrustration effects were not observed on other subscales, the results were interpreted as an instance of "targeting" of a minority group within the general framework of the scapegoat phenomenon.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. ALLPORT, G. W. *The nature of prejudice*. Cambridge: Addison-Wesley, 1954.
3. ALLPORT, G. W., & KRAMER, B. M. Some roots of prejudice. *J. Psychol.*, 1946, **22**, 9-39.
4. BETTELHEIM, B., & JANOWITZ, M. *Dynamics of prejudice*. New York: Harper, 1950.
5. CONGDON, C. S. Effects on attitudes resulting from experimental frustration. Unpublished master's thesis, Univer. of Illinois, 1954.
6. COWEN, E. L. The influence of varying degrees of psychological stress on problem-solving rigidity. *J. abnorm. soc. Psychol.*, 1952, **47**, 512-519.
7. DOLLARD, J., DOOB, L. W., MILLER, N. E., MOWRER, O. H., & SEARS, R. R. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.
8. FREUD, S. *The basic writings of Sigmund Freud*. New York: Random House, 1938.
9. GOUGH, H. G. Studies of social intolerance: I. Some psychological and sociological correlates of anti-Semitism. *J. soc. Psychol.*, 1951, **33**, 237-246.
10. HARDING, J., KUTNER, B., PROSHANSKY, H., & CHEIN, I. Prejudice and ethnic relations. In G. Lindzey (Ed.), *Handbook of social psychology*. Vol. II. Cambridge: Addison-Wesley, 1954. Pp. 1021-1061.
11. HOROWITZ, E. L. "Race" attitudes. In O. Klineberg (Ed.), *Characteristics of the American Negro*. New York: Harper, 1944. Pp. 139-247.
12. KATONA, G. *Organizing and memorizing*. New York: Columbia Univer. Press, 1940.
13. KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems of social psychology*. New York: McGraw-Hill, 1948.
14. LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
15. LINDZEY, G. An experimental examination of the scapegoat theory of prejudice. *J. abnorm. soc. Psychol.*, 1950, **45**, 296-309.
16. MILLER, N. E., & BUGELSKI, R. Minor studies in aggression: The influence of frustrations imposed by the in-group on attitudes expressed toward out-groups. *J. Psychol.*, 1948, **25**, 437-442.

17. MORSE, NANCY C., & ALLPORT, F. H. The causation of anti-Semitism: An investigation of seven hypotheses. *J. Psychol.*, 1952, **34**, 197-233.
18. MUSSEN, P. N. Some personality and social factors related to changes in children's attitudes towards Negroes. *J. abnorm. soc. Psychol.*, 1950, **45**, 423-441.
19. PASTORE, N. The role of arbitrariness in the frustration aggression hypothesis. *J. abnorm. soc. Psychol.*, 1952, **47**, 728-732.
20. ROSENBLITH, JUDY F. A replication of "Some roots of prejudice." *J. abnorm. soc. Psychol.*, 1949, **44**, 470-489.
21. SAPPENFIELD, B. R. *Personality dynamics*. New York: Knopf, 1954.
22. STAGNER, R., & CONGDON, C. S. Another failure to demonstrate displacement of aggression. *J. abnorm. soc. Psychol.*, 1955, **51**, 695-696.
23. ZAWADSKI, B. Limitations of the scapegoat theory of prejudice. *J. abnorm. soc. Psychol.*, 1948, **43**, 127-141.

Received September 3, 1957.

ACQUIRED AND SYMBOLIC AFFECTIVE VALUE AS DETERMINANTS OF SIZE ESTIMATION IN SCHIZOPHRENIC AND NORMAL SUBJECTS¹

THEODORE P. ZAHN²

Duke University

RECENT work in psychopathology has shown that differences between schizophrenics and normals, with respect to some kinds of performance at least, may be apparent primarily when stressful or threatening elements are present in the situation (14). Criticism of subject's (S's) performance (3, 12) and the use of stimulus materials depicting a censorious relationship between a mother and a son (9) have both been found to affect differentially the performance of schizophrenic and normal Ss in experimental tasks.

An important variable related to reactions to stress and threat in schizophrenics has been the level of social and sexual adjustment before the illness as measured by the Phillips Scale of Premorbid Adjustment in Schizophrenia (16). This scale was designed to measure the extent to which the patient had participated in interpersonal relationships before his illness. Patients with very inadequate premorbid adjustments (Poors) have been found to have less favorable prognoses than those with more adequate premorbid histories (Goods) (10, 16). These two kinds of schizophrenic patients have been found also to exhibit different reactions to stress in experimental situations. The Poors exhibited poorer performance in a discrimination task than did Goods or Normals only when pictorial representations of censorious mother-son relationships were involved (9).

¹ This report is based on a doctoral dissertation submitted to the Graduate School of Arts and Sciences of Duke University. The research was supported by USPHS-NIMH Project M-629. The author wishes to express his gratitude to Eliot H. Rodnick for advice and encouragement throughout all phases of the study and in the writing of this report. He is also indebted to the departments of psychology and psychiatry at the Roanoke VA Hospital, the Durham VA Hospital, and the State Hospital at Butner, N. C., for their cooperation in making subjects and facilities available. A paper based on the findings in the present report was read at the APA Convention in 1956, at Chicago, Illinois.

² Now at the Laboratory of Psychology, National Institute of Mental Health, Bethesda 14, Md.

A study by Bleke (3) also demonstrated that censure or threat is more effective in modifying the behavior of Poors than Goods, but that the modification may not necessarily be detrimental. These modifications seem to be in the form of behavior directed toward an avoidance of the censure or threat which may or may not interfere with effective task performance, depending on the situation (3, 4, 17).

Harris (13) had Good and Poor premorbid schizophrenics and normal Ss judge the sizes of a square, a "neutral" picture, and four pictures of a mother and son with various emotionally relevant contents. On the latter five items, the Poors overestimated and the Goods underestimated the sizes of these pictures, with the Normals falling in between. On the basis of Klein's (15) notion that Ss who manifest flexible or weak cognitive controls over their needs tend to overestimate the sizes of stimuli as compared to Ss whose control mechanisms are stronger or "constrictive," these data suggest that the two premorbid groups may differ in the degree of such control. Klein's data also show that the presence of a strong need will exaggerate the dominant size estimation tendency, although the evidence for this in the Klein study is convincing only for the flexible control Ss. Other studies on size estimation indicate that overestimation may be a function of either positive or negative value or "personal relevance" (5, 6, 8). Under this perceptual accentuation hypothesis, Harris' data may be explained by assuming that all the pictures he used had more personal relevance for the Poors than for the other two groups.

According to both the Klein hypothesis and the perceptual accentuation hypothesis, then, one would expect changes in size estimation as a function of the affective value of the stimulus being estimated. These theoretical and empirical considerations indicate that the size estimation technique may have promise as a

means of investigating the kinds of variables which have special affective significance for schizophrenic patients and for elucidating the mechanisms underlying their behavior in situations possessing such significance.

It was thought that affective value might be "built into" neutral pictures by a pseudo-conditioning technique involving differential reinforcement. This was attempted by prior presentation of the two to-be-judged pictures in a discrimination task *not* involving differential size. *S*'s responses to one of the pictures were almost consistently rewarded (he was informed that he was "right"). His responses to the other picture were just as consistently punished (informed "wrong"). Size judgments of each picture were then made by the method of average error.

It may be assumed that through the conditioning procedure, the punished picture acquires secondary reinforcing properties and produces to some extent the same effects as those elicited by "wrong." This may be functionally equivalent to an increase in a drive. Thus, from the Klein hypothesis, one would expect an increased overestimation of the censured picture by the Poors and an increased underestimation of it by the Goods. The perceptual accentuation hypothesis (in conjunction with the Poors' demonstrated greater sensitivity to censure) would lead similarly to the prediction of an increased overestimation by the Poors but would predict smaller changes in the same direction by Goods and normal *S*s. A preliminary study involving ten Poors and six Goods, in addition to confirming Harris' finding of a relative overestimation tendency by the Poors, showed that this group overestimated the size of the picture associated with punishment relative to the rewarded picture, and that the Goods showed a small tendency in the opposite direction. The interaction between premorbidly and type of reinforcement was statistically significant ($p < .05$).

The present experiment was designed partly to examine this phenomenon more definitely. Half of each subject group judged the sizes of two neutral pictures after the latter had affective value built into them by the discrimination procedure. The remaining *S*s judged these same pictures without their having been so treated. This latter procedure provides a test

of the hypothesis that the subject groups differ in their general size estimation tendencies to neutral stimuli.

Similarly, the sizes of two mother-son pictures, one a scolding scene and one a feeding scene, were judged both "treated" and "untreated." The hypothesis was entertained that the symbolic representation of a censorious mother-son relationship produces effects similar to those produced by the experimental administration of censure. Under the assumption that the Poors have been subjected most strongly to maternal domination (17), it was predicted that they overestimate the size of the scolding picture relative to that of the feeding picture to a greater extent than the other two groups.

In addition, the experiment may shed some light on the more general problem of the motivational conditions which influence size estimation. The present procedure of adjusting the size of the actual stimulus to a standard from immediate memory is not strictly comparable to the previous studies which used discs and had the standard present at all times, but it is probable that similar variables operate in both cases.

METHOD

Subjects

All *S*s were white, male hospitalized veterans, less than 45 years of age, without gross or uncorrected sensory defects or organic brain damage so far as could be determined, and of normal intelligence. Twenty-four of the *S*s were diagnosed as schizophrenic. Of these patients, 13 were rated as Goods and 11 as Poors on the basis of the Phillips Scale of Premorbid Adjustment in Schizophrenia. The range of possible scores on this scale is 0-30; a high numerical value indicates a poor premorbid adjustment and a low score a good premorbid adjustment. In the present study, the Poors had scores of 18 and above, the Goods 13 and below. Only those patients were used whose file contained sufficient information to permit a reliable rating. No patient was used who had had a lobotomy or who had undergone shock therapy for at least one month prior to testing. Most patients had had shock some time in the past. Although 16 patients taking tranquilizing drugs were tested, no data from such *S*s are included in the present report. Fifteen schizophrenics, 10 Poors and 5 Goods, were excluded from the sample because of inability to perform the task adequately.

Twenty normal *S*s were drawn from the medical and surgical wards of a VA general hospital. Case records were scrutinized and ward doctors and nurses consulted to choose patients who were free from major psychiatric symptoms. Patients whose diagnoses included disorders usually considered to be psychosomatic were

not used. No normals were eliminated because of failure to do the tasks adequately.

The Size Estimation Task

The apparatus for the size estimation task has been described in detail elsewhere (13). The stimuli were negatives (white lines on black background) projected by a 10-watt point source of light on a translucent screen large enough (22" x 32") to minimize framework cues. The distance of the stimulus slides from the light source (and hence the size of the image) was adjustable by either of two knobs. One knob was at the back of the apparatus and was for the use of *E*; the other protruded just below the screen and was operated by *S*, who sat in front of the screen. A plywood shield extended across the base of the screen and over *S*'s knob to shield it from his view. The use of a point source of light insured a focused image of constant brightness on the screen at all sizes. The image on the screen could be turned on and off by means of a shutter.

Changes in the position of the carriage in time were recorded by means of a pen in a holder hinged to a brass rod extending from the carriage to a gridded chart which was driven by a continuous feed kymograph.

The *S*s estimated the sizes of the pictures from immediate memory by the method of average error. On each trial, *E* said, "Here's the regular size," and flashed the picture on for about five seconds by count. Then he shut it off, changed the position of the carriage, and re-exposed the picture at a different size saying, "Now change it," as he did so. Approximately four seconds elapsed between exposures. When *S* had adjusted the picture to suit himself, he indicated so verbally, and the picture was shut off.

Each picture was judged eight times by this method. The size at which the picture was re-exposed (offset) varied from trial-to-trial in a predetermined order, constant for all *S*s on all pictures. There were four different offsets, two in each direction. For each of the first four trials and for each of the last four trials for any one picture, a different offset was used.

The Discrimination Task

In this task, each *S* was required to discriminate standard from variations for two pictures. The apparatus included two Kodaslide projectors, one of which had an Alphax shutter mounted on it, and a beaded screen. On any given trial, *E* exposed the standard on the screen for two seconds; this was followed by a two-second wait, and then the variable picture was presented for $\frac{1}{2}$ second. *S* indicated whether he thought the two pictures the same or different by pushing or pulling a Mallory switch which was appropriately labeled. *E* reinforced *S* by lighting up one of the two boxes, one labeled "right" and the other "wrong," placed on the floor just below the screen. The front side of each box consisted of a sheet of translucent plastic on the back of which was printed either "right" or "wrong" in black letters 1- $\frac{3}{4}$ " high. A 15-watt incandescent bulb was in each box. The words were faintly visible when the boxes were unlighted and stood out clearly when they were lighted.

Each *S* was given 30 trials on this task, 15 on each

of two pictures. Trials on the two pictures were mixed in a predetermined irregular order with not more than two consecutive trials on the same picture. For one of the pictures a discrimination was possible, the variable picture being different from the standard on two-thirds of the trials, and *S* was rewarded (flashed "right"), on 13 of the 15 trials, and punished on only two trials regardless of his responses. It was hoped that this discrimination would be fairly easy and that most *S*s would really master it. For the other picture, no discrimination was possible, the standard and variable stimuli being the same on all trials, and *S* was told "wrong" on 13 of the 15 trials and "right" on two trials regardless of his responses. These two pictures are referred to as the "rewarded" and "punished" pictures, respectively.

There were five stimulus pictures: a square and four pictures with meaningful content. Two of the content pictures contained objects: one, a house and a tree (H-T), the other a lamp and a table (L-T). The other two pictures depicted a woman and a small boy. In the "scolding" (Sc) picture, the woman had her arm outstretched, and the boy stood with his head down.³ In the "feeding" (Fe) picture, the woman held a pitcher outstretched and the boy a glass in his outstretched hand. The variations of these pictures used in the discrimination task were the same as the standard except in the positions of one element. In the house-tree picture, the angle of a limb varied; in the lamp-table picture, one leg was in different positions; on the mother-son pictures, the angle of the mother's arm varied. The pictures were designed to be structurally similar to a degree, all containing one tall and one short figure, and the two mother-son pictures were the most similar. The heights of the taller figures were equal.

Procedure

When *S* was seated before the size estimation apparatus in the experimental room, he was told that *E* was interested in how well people could judge the sizes of different pictures. The instructions described the sequence of events involved in each judgment, and *S* was urged to make the picture as close to the regular size as he could and told that he could take his time in doing so. *E* then darkened the room and gave *S* at least two practice trials with the square, reiterating parts of the instructions during this process. If *S*'s two practice adjustments met a predetermined criterion of accuracy, a regular series of trials was begun. If *S*'s two practice adjustments were inaccurate, more practice was given until he improved sufficiently or until it became clear that his performance would probably never be valid. Each *S* then judged the sizes of the square and one content picture, then was given the discrimination task, and finally judged the two pictures seen in the discrimination task and one other.

The following descriptions were given of the pictures just before *S* was presented with them for the first time, whether in the size estimation or discrimination tasks: H-T: "Here is a picture of a house and a tree." L-T: "Here is a picture of a lamp and a table." Sc: "Here is a

³ The H-T and Sc pictures are based on those used by Dunn (9).

picture of a mother scolding her little boy. He has done something bad and she is bawling him out for it." Fe: "Here is a picture of a mother and her little boy. She's about to pour him a glass of milk out of the pitcher." At the end of the session, *S* was highly praised for his performance on both the size estimation and discrimination tasks. His failure on the punished picture was discounted as "just one of those things," and he was told that it was a very tricky picture anyway and that many people did not do well on it.

Experimental Design and Treatment of Data

Each main subject group was divided into four subgroups, differing in the order in which the pictures were presented and in which pictures were used in the discrimination task. Table 1 shows this sequence and which pictures were judged after having been rewarded (R) and punished (P). This design was identical for each subject group. The *N* for each condition for the schizophrenics is 3 except for the Goods in Condition 1, where *N* = 4, and the Poors in Condition 4, where *N* = 2. For the Normals, *N* = 5 for each subgroup.

The measure of size estimation analyzed here is the distance of the slide from the standard position, hence the distance of the slide from the screen. This measure was used in order to approximate an equal discriminability function. The relationship between this measure and the logarithm of the height of the image on the screen closely approaches linearity. The score for any given *S* is the mean of 8 of these values, except where otherwise indicated. A positive value refers to a judgment larger than the standard size, and a negative value to a smaller estimation.

RESULTS

Reward-Punishment Treatments

Group means of the size estimation means of the rewarded and punished pictures for all conditions and for total are presented in Table 2. It is evident that, contrary to expectation, all groups show a general underestimation tendency except for the Goods on the punished picture. Group differences were evaluated statistically by analyzing the difference between the means of the punished and rewarded pictures for each *S*. The breakdown included first four vs. last four trials, neutral vs. mother-son pictures, and order of

TABLE 1
EXPERIMENTAL DESIGN: IDENTICAL
FOR EACH SUBJECT GROUP

Con- dition	Sequence of Pictures Judged			
1	Fe	Rwd. (H-T)	Pun. (L-T)	Sc
2	Sc	Pun. (H-T)	Rwd. (L-T)	Fe
3	H-T	Rwd. (Fe)	Pun. (Sc)	L-T
4	L-T	Pun. (Fe)	Rwd. (Sc)	H-T

TABLE 2
SIZE ESTIMATION MEANS IN TERMS OF DEVIATIONS
FROM STANDARD SIZE FOR THE VARIOUS
GROUPS AND CONDITIONS

Condition	Good Premor- bid Schiz.		Poor Premor- bid Schiz.		Normals	
	Rwd.	Pun.	Rwd.	Pun.	Rwd.	Pun.
Rwd. on H-T						
Rwd. Judged First	-.192	.160	-.102	-.323	-.154	-.048
Rwd. on L-T						
Rwd. Judged Second	-.054	.108	-.090	-.027	-.153	-.186
Subtotal	-.133	.139	-.096	-.175	-.154	-.117
Rwd. on Fe						
Rwd. Judged First	-.521	.258	-.273	-.083	-.294	-.234
Rwd. on Sc						
Rwd. Judged Second	.048	.104	-.219	-.460	-.300	-.254
Subtotal	-.236	-.077	-.251	-.234	-.297	-.244
Total	-.181	.039	-.166	-.202	-.226	-.180

estimation of the two pictures as well as groups. An analysis of variance, a modification by the addition of a fourth dimension of Lindquist's Type III design, was carried out on these difference scores after Bartlett's test had revealed no significant heterogeneity of variance.⁴ The variance attributable to groups was significant ($F = 3.64$, $df = 2, 32$, $p < .05$), and *t* tests of differences between individual means show that the Goods differed significantly from the Poors ($p < .05$), the Goods-Normals difference approached significance ($p < .10$), and the Poors and Normals did not reliably differ ($p > .20$).

The hypothesis of a difference between the reward and punishment treatments was tested for each of the groups by comparing the *P-R* differences to zero. It was found that the Goods significantly ($p < .01$) judged the punished picture larger than they did the rewarded one, but the differences between the two treatments for the Poors and Normals were not significant. It is clear that the significant difference between groups comes from the differential impact of the reward and

⁴ To meet the requirement of proportionate cell frequencies on this and subsequent analyses, the score closest to the mean was eliminated from the larger cell and the mean of the smaller cell added to that cell. This latter procedure necessitated the subtraction of two *df* from total and from error (*w*) in the analysis of variance.

punishment conditions on the Goods as contrasted with no appreciable effects in the other two groups. The only other source of variation in the table to reach significance was the trials \times order interaction ($p < .05$). This reflects a tendency for all Ss to underestimate rather greatly on the first trial of the first picture after the discrimination task.

The differences between the R and P conditions were, in general, less on the last four trials than on the first four. On the first four trials, in fact, the Normals showed a significant ($p < .05$) overestimation of the punished picture relative to the rewarded one, as did the Goods. Although, in view of insignificant interactions involving trials in the larger analysis, this result can be looked on only as suggestive, it indicates that the reward-punishment variable may have had a transitory effect on the size estimations of the Normals, but that this effect dissipated in this group more rapidly than for the Goods.

Feeding and Scolding Content

Mean size estimation scores for the Feeding and Scolding pictures for the subgroups which were *not* given the reward-punishment treatments on these pictures are shown in Table 3. The means are strikingly similar to those for the treatments, showing a general underestimation tendency holding for all but the Goods on the Scolding picture. Means for the subgroups which were "treated" on the mother-son pictures are also shown, as well as means for all four conditions. A $3 \times 4 \times 2$ Type III analysis of variance on the means for all four conditions shows that the groups variance approaches significance ($F = 3.13$, $df = 2$, 32 , $p < .07$). Examination of Table 2 shows that this lack of significance may well be due to the confounding effects of the punishment on the feeding picture on the last condition. An analysis of the means for only those conditions in which the mother-son pictures were *not* involved in the treatments would seem to provide a less contaminated test of the hypothesis. Consequently, a $3 \times 2 \times 2$ Type III analysis of variance was carried out on these untreated means. The variance attributable to differences between groups is significant ($F = 3.69$, $df = 2$, 16 , $p < .05$), indicating that the content of the pictures has a differential effect on these three groups.

TABLE 3
MEAN SIZE ESTIMATES OF THE FEEDING AND SCOLDING PICTURES WHEN NOT SUBJECTED TO TREATMENTS

Order	Good Premorbid Schizophrenics		Poor Premorbid Schizophrenics		Normals	
	Fe	Sc	Fe	Sc	Fe	Sc
Fe-Sc	-.397	-.008	-.192	-.323	-.198	-.121
Sc-Fe	-.021	.113	-.139	-.019	-.095	-.246
Mean	-.236	.044	-.166	-.171	-.146	-.184
Mean (Treated Subgroups)	-.208	-.105	-.348	-.137	-.274	-.267
Grand mean	-.223	-.025	-.249	-.156	-.210	-.226

Differences between the separate groups were tested by t tests, which resulted in the findings that the difference between the two schizophrenic groups approaches significance ($t = 2.09$, $p < .06$), the Goods differ significantly from the Normals ($t = 2.61$, $p < .05$), and the Poores and Normals are not significantly different. Here again it is the Goods who manifest a reaction to the experimental variables.

This analysis was based on differences between the picture judged second in the series and the one judged last (fifth). Because the intervening experience with the reward and punishment treatments may have influenced, by some kind of carry-over effect, the size judgments of the last picture, the judgments of the second picture only were studied. A type III analysis of variance, including groups, pictures, and trials, showed a significant groups \times pictures interaction ($F = 3.73$, $df = 2$, 16 , $p < .05$), which confirms the results of the other analysis. Individual t tests show that the Goods were the only group to judge the pictures differently from one another ($t = 4.48$, $p < .005$).

An examination of Conditions 3 and 4, where the treatments were given on the mother-son pictures, indicates something of the interaction between the two kinds of affective meaning. For the Goods on Condition 3, where the treatments and content are congruent, the difference between the size means for the two pictures is not markedly different from the conditions where the variables operated singly. This indicates the absence of any summative process that might be expected. On the fourth condition where treatments and content are incongruous, the difference be-

tween pictures is much the smallest of the four conditions, indicating that the two sources of negative affective meaning cancelled each other out. The difference between differences on these two conditions is not significant ($p < .11$), however. The fact that both pictures are rather extremely underestimated on Condition 3 and slightly overestimated on Condition 4 cannot be stressed since these same Ss showed somewhat comparable effects on the neutral pictures.

The data for the Poors reveals that for both these conditions the Scolding picture is underestimated much less than Feeding, a difference significant at the .05 level by t test. Previous experience with these pictures in the discrimination task, regardless of its nature, may have enhanced the symbolic value of their content for the Poors.

Neutral Pictures

The size judgments of the square and of the untreated neutral pictures were analyzed to test the hypothesis of differential over-all size estimation tendencies. In neither case do the differences between subject groups approach significance, although the Poors show a smaller underestimation trend on the untreated neutral pictures.⁵ In addition, there were no marked differences between the two neutral pictures for any group, despite relatively marked differences in structural properties. This indicates that the differences between the two mother-son pictures demonstrated above is probably not due to such structural factors.

Other Aspects of the Adjustment Process

Another important aspect of the method of average error is the consistency with which the judgments are made. This is given by the variability of a given set of individual judgments around their mean (variable error). This statistic was computed for each S on each picture. The over-all means of this measure for all Ss on all pictures are .459 for the Goods, .464 for the Poors, and .482 for the normals. It is interesting to note that these over-all differences are very small and that variability is slightly greater for the normals

than for the schizophrenics. No consistent relationships between variability and either the size estimation mean or the affective value of the stimulus could be demonstrated.

Two kinds of data were quantified from the graphic records of the Ss' adjustments: the time taken in making the adjustments, and changes in the direction of adjustment. For the time of adjustment, the over-all averages show that the Goods took an average of 10.07 seconds per trial, the Normals 7.50 seconds, and the Poors 7.68 seconds.⁶ A simple analysis of variance shows the difference significant ($F = 3.93$, $p = .05$, $df = 2/39$) and that the Goods differ from both the other groups.

A change in the direction of adjustment refers to the cases where S overshot the mark and moved the image back in the opposite direction from the way in which he had been moving it. It was thought that this, as well as the time variable, might reflect a hesitancy in coming to a decision. The mean number of trials per picture on which one or more of these direction changes occurred is 3.01 for the Goods, 2.28 for the Poors, and 1.24 for the Normals. The three groups were significantly different ($F = 6.68$, $p = .005$, $df = 2/40$). The goods differed from the Normals ($p < .001$), but the Poors were only suggestively different from the Normals ($p < .10$) and not different from the Goods. For both these variables, the differences between pictures for the three groups were quite small and nonsignificant.

DISCUSSION

The results on size estimation, although they are contradictory of the specific research hypotheses that were advanced, confirm the more general hypothesis and research finding that the two schizophrenic groups differ in reactions to stressful or threatening situations (3, 13, 17).

In addition, the experiment demonstrates that having been consistently censured in responding to a stimulus is a variable influencing the estimation, from immediate memory, of the size of the stimulus, at least for one class of S . It also suggests a functional equivalence

⁵ The means for the Goods, Poors, and Normals for the untreated neutral pictures are $-.228$, $-.078$, and $-.267$, respectively (compare with Tables 2 and 3).

⁶ The mean for the Poors omits one extreme S whose time of adjustment was over twice as long as that of the next most extreme case. This was done on the basis of a test for extreme scores (7, p. 243).

between the experimental application of censure and the symbolic representation of a censorious mother-son relationship.

In view of the failure to find consistent differences in over-all size estimation tendencies for the three groups, Klein's control hypothesis seems inapplicable to the results of the present study. According to the accentuation hypothesis, however, the interpretation appears at least tentatively justifiable that a reliable relative overestimation in the present experiment indicates the picture so judged had more personal meaning for the Ss than had pictures not so judged. The mechanisms by which an attitude is translated into a change in estimated size remain obscure.

A question of more primary concern here is what mechanisms underlie increased personal relevance. The discrimination task was designed to present the Punished picture in close conjunction with a negative reinforcement. Thus, the picture would be expected to acquire secondary reinforcing properties and to evoke the same kind of reaction as is produced by the punishment itself. This reaction is probably emotional in nature. The magnitude of this emotional reaction produced by the punished picture is probably greater in persons characterized as having a high degree of anxiety or affect. These considerations lead to the hypothesis that the Goods used in the present experiment are characterized by a higher degree of anxiety or affective sensitivity than are the Poors.

The same kind of reasoning may be applied to the mother-son pictures. Here the emotional reactions to situations of the kind pictured by the Scolding picture are assumed to have already been conditioned by virtue of the harsh treatment schizophrenics are thought to have received from their parents. The picture, then, presumably serves as a cue to evoke, at least partially, the same kind of reaction as was engendered by the situation it represents. The strength of this reaction is probably a function both of the experience of S with his mother and of the amount of anxiety or affect characteristic of him. It is interesting to note that the Goods gave evidence of as great a change in size estimation to the Scolding picture as to the Punished picture. Although the Poors were relatively unresponsive to the treatment variable (in terms of size estimation changes),

they showed evidence of being sensitive to the mother-son pictures under some conditions.

The relatively long time of adjustment and large number of direction changes found for the Goods may indicate a greater motivation to succeed in a task. This motivation may, in turn, be functionally related to anxiety.

The results for the patients taking tranquilizing drugs seemed to provide some evidence for the anxiety hypothesis. It was found that the differences in the Goods from the other groups were not present in the patients taking drugs (18). Previous clinical and experimental studies on the effects of chlorpromazine on psychiatric patients have noted a lessening of clinically manifested and subjectively felt anxiety (1, 11) and pathological affect (19) after a period of administration of the drug. It seems reasonable, then, to assume that the dependent variables in the present experiment on which differences occurred between no-drug and chlorpromazine Ss may be influenced by motivational states such as anxiety. In other words, the fact that the effects of the punishment and Scolding content were absent for those patients taking chlorpromazine indicates that the effects observed in the no-drug groups were a function of a relatively high state of anxiety or affective responsivity. However, this argument is weakened by lack of control over the sampling of the patients taking drugs.

The results presented here conflict with previous studies which show greater reactions to stressful and threatening situations in poor premorbid schizophrenics than in Goods or Normals, and specifically with the results of Harris' study (13) and the preliminary experiment which showed a relative general overestimation tendency by the Poors. The experimental situation of the present study, although somewhat different from that of Harris, was quite similar to that of the preliminary study. This makes it somewhat unlikely that differences in technique are responsible for the conflicting results of the present study, although this possibility cannot be entirely ruled out. It would seem more probable that the difference is a function of the sampling of the Ss. One striking difference stands out in this respect—the mean length of time from the first psychiatric hospitalization of the patients. In the present experiment, this was 6.00 years for the Goods and 8.09 years for the Poors, as

compared with about three years for Harris' total group and less than that for the Ss used in the preliminary experiment. Thus, it can be inferred that the patients in the present study were drawn from a different population as far as this chronicity variable goes.

There seem to be two ways in which this chronicity variable could work to produce a different sampling of patients. First, there may be some effect of chronicity per se on behavior in the size estimation task. Second, sampling from a chronic population would lead to choosing patients with poor prognoses and miss patients who have improved and left the hospital. Also, more formerly available patients would have deteriorated to the point of unsuitability as experimental Ss.

It is clear that the Goods in the present study were atypical from the standpoint of prognosis since this kind of patient has been found to improve more readily than Poors (10, 16). It may be speculated that their sensitivity to stress manifested here has had some relation to their demonstrated lack of improvement. The sensitivity to the Scolding picture may indicate that these patients had a more typical Poor premorbid familial pattern.

For the Poors, on the other hand, the gradual progression of the disease may be responsible for the lack of effect of the motivational variables on size estimation. A concomitant of long hospitalization in chronic schizophrenics is a decreased sensitivity to the external and internal environment, usually described as withdrawal, decreased motivation, or increased apathy (2, 14). Such a general decrease in affective responsivity may be responsible for the negative findings for the Poors. The two years' difference in hospitalization time would seem not to be enough, however, to account for the differences between the two premorbid groups. It is possible that the difference in premorbidness may manifest itself (in patients who do not leave the hospital) in differential rates of progression of the illness.

However, these speculations on the effects of hospitalization length are not confirmed by the present data. No consistent differences in performance can be found in either group by separation into subgroups on the basis of length of hospitalization. This cannot be regarded as a crucial test, however, since even the patients in the short hospitalization groups

in this breakdown were psychotic for a longer time than Ss used in previous studies.

The possibility cannot be ignored that the Poors, in view of their chronicity, were completely "out of" the tasks and behaved in a more or less random manner. However, this seems unlikely since they were comparable to the other groups in terms of the intra-individual variability of the size judgments and in the magnitudes of the constant errors on most of the pictures. Also, in the discrimination task on the punished picture, they gave significantly more reversals of response after a "wrong." These reversals have been interpreted by Rodnick and Garmezy (17) as indicating attempts by *S* to avoid punishment. This suggests that the punishment had *some* effect on the behavior of the Poors.⁸ The failure of this effect to manifest itself in the size estimations of the Poors may itself be a consequence of strong avoidance tendencies in these Ss. In view of their long history of avoidance of and withdrawal from interpersonal situations, it may be speculated that they were able to avoid the affective properties of the stimuli; that is, they were able to avoid getting involved in the size estimation task.

SUMMARY

The present experiment investigated the size estimations from immediate memory of pictures imbued with two kinds of affective meaning by schizophrenics with good premorbid adjustments, schizophrenics with poor premorbid adjustments, and normals. Positive and negative affective meanings were built into two pictures by means of a prior task involving differential reinforcement (right and wrong) of *S*'s responses to them. In addition, size estimates were made of pictures depicting a mother scolding a boy and feeding a boy. The good premorbid schizophrenics, in contrast to the other two groups, significantly overestimated the sizes of both the picture associated with "wrong" and the Scolding picture, relative to the rewarded and Feeding pictures. The results are interpreted in terms of a high degree of anxiety or affective responsivity in

⁷ A reversal is said to occur if *S* responds "same" on one trial and "different" on the next or vice versa.

⁸ There were no significant differences between the three groups in number of correct responses to the rewarded picture.

the Goods and the predominance of avoidance and withdrawal mechanisms in the Poor. They also demonstrate the possible fruitfulness of the size estimation technique in discovering affective variables influencing the behavior of schizophrenic patients.

REFERENCES

1. ANTON-STEPHENS, D. Preliminary observations on the psychiatric uses of chlorpromazine (Largactil). *J. ment. Sci.*, 1954, **100**, 453-457.
2. AREITH, S. *Interpretation of schizophrenia*. New York: Brunner, 1955.
3. BLEKE, R. Reward and punishment as determiners of reminiscence effects in schizophrenic and normal subjects. *J. Pers.*, 1955, **23**, 479-498.
4. BLEKE, R. Interfering response sets as determining factors of reminiscence effects in schizophrenic subjects. *Amer. Psychologist*, 1954, **9**, 334. (Abstract)
5. BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factors in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33-44.
6. BRUNER, J. S., & POSTMAN, L. Symbolic value as an organizing factor in perception. *J. Soc. Psychol.*, 1948, **27**, 203-208.
7. DIXON, W. J., & MASSEY, F. J. *Introduction to statistical analysis*. New York: McGraw-Hill, 1951.
8. DUKES, W. S., & BEVAN, W. Size estimation and monetary value, a correlation. *J. Psychol.*, 1952, **34**, 43-53.
9. DUNN, W. L. Visual discrimination of schizophrenic subjects as a function of stimulus meaning. *J. Pers.*, 1954, **23**, 48-64.
10. FARINA, A., & WEBB, W. W. Premorbid adjustment and subsequent discharge. *J. nerv. ment. Dis.*, in press.
11. FROMM, G. H., & FORSBERG, I. A. Controlled study on the value of chlorpromazine in allaying anxiety. *Dis. nerv. Syst.*, 1956, **17**, 16-19.
12. GARMEZY, N. Stimulus differentiation by schizophrenic and normal subjects under conditions of reward and punishment. *J. Pers.*, 1952, **21**, 253-276.
13. HARRIS, J. G., JR. Size estimation of pictures as a function of thematic content for schizophrenic and normal subjects. *J. Pers.*, in press.
14. HOSKINS, R. G. *The biology of schizophrenia*. New York: Norton, 1946.
15. KLEIN, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Vol. II. Lincoln: Nebraska Univer. Press, 1954. Pp. 224-274.
16. PHILLIPS, L. Case history data and prognosis in schizophrenia. *J. nerv. ment. Dis.*, 1953, **117**, 515-525.
17. RODNICK, E. H., & GARMEZY, N. An experimental approach to the study of motivation in schizophrenia. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Vol. V. Lincoln: Univer. of Nebraska Press, 1957. Pp. 109-184.
18. ZAHN, T. P. Acquired and symbolic affective value as determinants of size estimation in schizophrenic and normal subjects. Unpublished doctoral dissertation, Duke Univer., 1956.
19. ZELLER, W. W., GRAFFAGNINE, P. N., CULLEN, C. F., & REITMAN, H. J. Use of chlorpromazine and reserpine in the treatment of emotional disorders. *J. Amer. med. Assn.*, 1956, **160**, 179-184.

Received September 10, 1957.

MENTAL ILLNESS, MILIEU THERAPY, AND SOCIAL ORGANIZATION IN WARD GROUPS

EDWARD J. MURRAY AND MELVIN COHEN¹

Waller Reed Army Institute of Research

CULTURALLY oriented personality theorists such as Sullivan have suggested that mental illness, broadly conceived, is related to difficulties in interpersonal relations (6). Fromm-Reichmann (2) views mental illness as a condition involving a withdrawal from social relationships because of fears of rejection and feelings of inadequacy which can be alleviated by a therapeutic process involving insight, emotional discharge, and respect by the therapist. This general theory of mental illness would lead one to predict that the greater the degree of mental illness of a patient the more his social relationships with his fellow patients in a mental hospital will show various disturbances such as withdrawal. McMillan and Silverberg (3) have presented evidence which generally supports this prediction.

A continuum of degree of mental illness was obtained by McMillan and Silverberg by ranking five wards according to decreasing level of adjustment: (a) neurological, (b) gastrointestinal medical, (c) open psychiatric, with neurotics and psychotics, (d) insulin therapy, with anxiety neurotics, and (e) closed psychiatric, with psychotics. Sociometric choice patterns were then compared with three results. First, there was a trend suggesting that the greater the degree of mental illness, the fewer reciprocal choices, which was, however, significant only for the insulin ward. Second, there was a slight but not significant trend suggesting that the more disturbed patients concentrated their negative choices on fewer people. Third, there was a striking tendency for patients on all wards to have more positive than

negative reciprocations. While these results generally tend to support the hypothesis, they are not all as clear and consistent as one would expect.

The purpose of the present study was to apply the McMillan and Silverberg technique to a somewhat more easily ranked group of wards and to employ somewhat simpler measures in the hope of obtaining a more clear-cut test of the hypothesis that the greater the degree of mental illness, the greater the disturbance of social relationships. An additional purpose was to take an initial step in determining the effect of milieu therapy on group organization in psychiatric wards.

METHOD

Major Ward Groups

A total of 132 patients in three wards at a large army hospital was used in the main part of the study. Nearly all of the patients were enlisted men. The *control medical ward* was a large orthopedic ward with a total of 40 patients, 5 of whom were omitted from the study because of absence or language difficulty. The majority of these patients had limb fractures or amputations but were ambulatory and able to socialize. The ward could be conveniently divided into front, back, and porch areas. The *open psychiatric ward* contained 51 patients, of whom 2 were omitted and 15 refused to participate. The diagnosis of the patients included various neuroses, nonacute psychoses, and character disorders. The men in this ward were considered by the staff to be almost well-adjusted enough to be returned to duty or discharged to their own care. If they became disturbed they were returned to the locked ward. The men had ground privileges and frequent weekend passes. There were common eating and recreational facilities, but the men slept in three separate areas. The *locked psychiatric ward* contained 41 patients, 13 of whom were omitted because of language difficulties or because they were not in contact enough to fill in their own names and identifying data on the questionnaire. This ward contained more acute psychotics and severe neurotics as well as patients who had not been fully diagnosed. The staff felt that these patients were considerably more disturbed than the patients in the open psychiatric ward. The patients were locked in the ward and escorted to all activities and examinations. They were not allowed passes. The ward had three sleeping areas as well as a central day room.

¹ The authors wish to thank Stanley Pliskoff, who helped with some of the statistical analysis, and Harold L. Williams and Ardie Lubin for reading the manuscript and making valuable suggestions. The authors also wish to thank the numerous staff members and patients of Walter Reed Army Hospital for their cooperation in this study. A portion of this study was presented at the American Psychological Association meeting in New York in August, 1957. The senior author is now at Syracuse University and M. Cohen is at New York University.

Additional Ward Groups

In addition to the three wards used in the main part of the study, three smaller wards were also used to study the effects of milieu therapy on group organization. The *nonpsychiatric control ward* was an experimental ward used for long-term studies of metabolic diseases. At the time of the initial sociometric questionnaire administration, there were eight patients on the ward. These patients were ambulatory but were restricted to the ward in order to control diet rigidly and to measure bodily waste materials. The *milieu therapy ward* was an experimental psychiatric ward for the long-term treatment and study of schizophrenia. The patients all had been hospitalized during basic training. On the ward, the patients were put in a therapeutically oriented milieu (4), including group therapy three times a week, varying hours of individual therapy with the psychiatrist, and frequent therapeutically oriented contacts with the nurse, social workers, and psychiatric aides all coordinated by weekly staff conferences. The staff was especially selected for empathic attitudes and high motivation. None of the patients received drugs, shock, or other somatic therapy. There were nine patients on the ward at the time of the first administration of the sociometric questionnaire. The *somatic therapy ward* was a typical closed psychiatric ward for disturbed patients. The patients were referred from smaller army hospitals and showed a greater age, rank, and diagnostic range than the ward just described. However, most of the patients were schizophrenic and all were confined to the locked ward. The patients were being treated with a long series of electroconvulsive shocks, a series of insulin comas, or massive daily doses of tranquilizing drugs. Very little psychotherapy was employed. The staff was competent and pleasant but not oriented towards milieu therapy. There were 15 patients on the ward at the time of the first sociometric questionnaire.

From four to eight months after the first administration of the sociometric questionnaire, the questionnaire was given again to the same three wards. By this time, all of the patients on the milieu therapy ward and the somatic therapy ward and all but three of the patients on the control ward had been replaced by new patients. At this time there were seven patients on the control ward, six on the milieu therapy ward, and eight on the somatic therapy ward. A few patients on all wards were omitted both times because they were on the ward for less than a week. The average length of stay was about equal on all wards.

Sociometric Questionnaire

The questionnaire was mimeographed in a booklet clearly marked confidential. The patients were asked to give their name, ward number, and other identifying data and were omitted if they could not reasonably fill out this section. Then, the patients were asked simply to list all of the names of the patients on the ward that they could remember. First names and nicknames were accepted as well as last names. Finally, the patients were asked to make choices on twelve sociometric items. They were asked to select at least

one person for each item but allowed to write in as many more as they wished. There were four spaces for names under each item but the patients were allowed to write in more if they wished to. The sociometric questionnaire included: four positive items—like most, at ease with most, like to eat with most, like to work with most; four negative items—like least, ill-at-ease with, like to eat with least, like to work with least; and four neutral items—tallest patients, heaviest patients, most intelligent patients, and best-liked patients. Later evidence suggested that the neutral items, particularly the last two, were influenced by the patients' likes and dislikes. The 12 sociometric items were mimeographed in a random order.

Procedure

Arrangements were made with the staff ahead of time to insure maximum attendance at a group administration of the questionnaire on each ward. The senior author explained the general nature of the questionnaire and emphasized its confidential nature. The booklet and pencils were handed out, and the patients worked at their beds or in the day room. Communications between the patients were kept to a minimum. All patients who were absent at the group meeting or who had refused to fill out the questionnaire were contacted individually and asked again to cooperate.

Social Background Factors

After the administration of the sociometric questionnaire, the hospital records of each patient were examined and information obtained about age, race, rank, length of service, marital status, education, population area of origin, branch of service, and religion. Information was also obtained about the length of time in the hospital and on the ward, diagnosis, and ward bed assignment.

RESULTS AND DISCUSSION

Degree of Mental Illness and Sociometric Choice

A number of sociometric measures were directly related to the degree of mental illness represented by the three major wards. This can be seen graphically in Fig. 1 where the sociograms based on positive reciprocal choices are shown for the control medical, open psychiatric, and locked psychiatric wards. Reciprocals are based on a mutual positive choice regardless of which of the four items was involved. These three wards form a continuum of degree of mental illness about which there can be very little disagreement. The sociograms indicate that as degree of mental illness increases there is a decrease in the complexity of the social organization as well as a decrease in the number of reciprocals. As Table 1 shows,

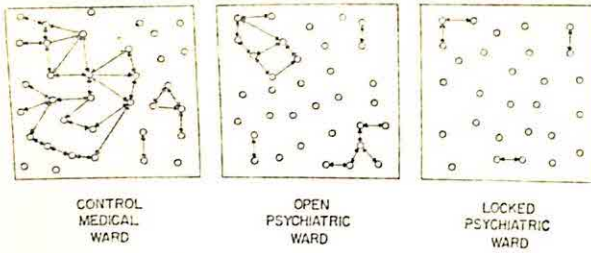


FIG. 1. POSITIVE RECIPROCAL CHOICES FOR THREE WARD GROUPS REPRESENTING A CONTINUUM OF DEGREE OF MENTAL ILLNESS

the percentage of patients with one or more reciprocals shows a significant variation over all wards ($p < .01$). (Unless otherwise stated, significance levels in this study are based on a one-tail chi-square test.) There are fairly significant decreases from the control medical to the open psychiatric ($p < .07$) and from the open psychiatric to the locked psychiatric ward ($p < .05$). The percentage of patients who had one or more negative reciprocals was significantly lower than for positive reciprocals for all wards ($p < .01$). However, the same trend as with positive reciprocals was evident with fewer people with negative reciprocals as degree of mental illness increased ($p < .02$). This suggests a general withdrawal rather than a differential effect of mental illness on positive and negative aspects of interpersonal relationships.

It is of interest to examine the number of reciprocal choices as well as the number of patients involved, since the number of reciprocal choices may reflect the complexity of the social organization more faithfully. For example, there may be a reciprocal choice between A and B and between B and C involving three patients altogether. However, this is a simpler organization than if there were also a reciprocal choice between A and C, which would increase the number of reciprocals but not the number of patients with one or more reciprocals. But in making such a comparison, it must be kept in mind that generalizations are limited to a population of not fully independent choices. It is felt that in this case such a comparison is useful and legitimate since the finding has already been established for a population of patients. The percentages of both positive and negative choices which were reciprocated in the control medical, open psychiatric, and

TABLE 1
VARIOUS SOCIOMETRIC MEASURES IN WARDS REPRESENTING A CONTINUUM OF DEGREE OF MENTAL ILLNESS

Sociometric Measure	Degree of Mental Illness		
	Control Medical Ward	Open Psychiatric Ward	Locked Psychiatric Ward
Percentage of patients with one or more positive reciprocal	74	50	25
Percentage of patients with one or more negative reciprocal	26	6	0
Percentage of positive choices reciprocated	44	38	22
Percentage of negative choices reciprocated	6	4	0
Average number of names known	14.2	8.8	7.2
Average number of positive choices	4.1	2.8	2.3
Average number of negative choices	2.5	1.5	1.0
Average number of neutral choices	4.3	2.2	2.2
Percentage of patients receiving no choices at all (total isolates)	0	18	21

locked psychiatric wards are shown in Table 1. As degree of mental illness increases there are fewer reciprocal choices on either positive or negative items. The over-all decrease of the percentages of positive reciprocals was significant ($p < .001$). The difference between the locked psychiatric and the open psychiatric wards was significant ($p < .01$) as was the difference between the locked psychiatric and the control medical wards ($p < .001$), but the difference between the open psychiatric and the control medical wards failed to reach significance on this measure. The percentages of the negative reciprocals showed an over-all decrease with degree of maladjustment ($p < .01$). The over-all difference between percentage of positive and negative reciprocals is significant ($p < .001$) but appears to be about equal for all wards.

Another way of showing this relationship is to simply compare the average number of names known on each ward. Table 1 makes this comparison; the more disturbed the ward, the fewer the names known. The over-all difference is significant ($p < .01$) as well as the difference between the control medical ward and the open psychiatric ward ($p < .05$) and the difference between the control medical

ward and the locked psychiatric ward ($p < .01$). The difference between the open and locked wards did not reach significance.

Similarly, the average absolute number of choices on the sociometric items decreases with degree of mental illness. Table 1 shows over-all decreases in average positive, negative, and neutral choices which are all significant ($p < .05$). The over-all average number of negative choices is significantly less ($p < .01$) than the average number of positive or neutral choices.

The number of social isolates increases with degree of maladjustment (Table 1). There is a significant difference in percentage of patients receiving no choices between the control medical ward and the two psychiatric wards combined ($p < .01$). The difference is accounted for almost entirely by a difference in the number of isolates on positive sociometric items, since from 40 to 60% of patients on all wards had no negative or neutral choices, a value that is significantly higher ($p < .001$) than the 20% who had no positive choices.

An analysis of variance showed that the three groups did not differ significantly in average length of time on the ward or in the hospital. Furthermore, the various measures showed almost zero correlations with number of weeks on the ward or in the hospital. This result would indicate that the decrease in interpersonal relationships from control medical to open psychiatric to locked psychiatric wards was not due to different opportunities for socialization.

It is also of interest to examine the degree to which the same person was chosen on several positive items or negative items. A score analogous to the criteria overlap score used by McMillan and Silverberg was obtained for each patient. There was no significant difference between the three ward groups on either positive or negative items. However, there was an over-all tendency for greater overlap on positive than negative items ($p < .01$). These results are in agreement with those of McMillan and Silverberg.

An acceptability score was also adapted from McMillan and Silverberg, consisting of the percentages of patients from all wards who were chosen various numbers of times on positive, negative, or neutral items. As would be expected, most patients were chosen very few

TABLE 2
PERCENTAGE OF SOCIOMETRIC CHOICES FOR
INDIVIDUALS SIMILAR TO SELF IN SOCIAL
BACKGROUND VARIABLES

Social Background Variable	Degree of Mental Illness		
	Control Medical Ward	Open Psychiatric Ward	Locked Psychiatric Ward
Age (within ± 5 years)	65	64	46
Race (Caucasian vs. Negro)	91	92	72
Population area (metropolitan vs. other)	63	71	47
Rank (Sgt. vs. Cpl. or Pvt.)	66	61	50
Length of service (under or over 3 years)	60	55	50
Marital Status (single vs. married)	61	56	44
Education (under or over ninth grade)	50	50	44
Geographical area (north vs. south)	55	68	56
Service (Army vs. Air Force)	68	52	64
Religion (Prot. vs. Catholic)	56	59	59

times if at all on any of the items, while a few people were chosen very often. But it was also found that fewer people received no choices at all on positive items than negative or neutral items. This result is contrary to McMillan and Silverberg's finding of fewer patients with zero neutral choices. Other differences between positive, negative, and neutral choices were not significant. McMillan and Silverberg's finding that more people had two positive choices than two negative or two neutral choices was not confirmed. Except for the differences already noted, there were no significant differences between wards.

Sociometric Choice and Similarity of Social Background

Another way of getting at the relationship between mental illness and interpersonal relations is to examine the extent to which patients on different wards choose fellow patients with social backgrounds similar to themselves. A number of social background variables that might be expected to influence sociometric choices are listed in Table 2. Each variable is dichotomized at some convenient point so that a 2×2 table could be set up to measure the degree to which patients made choices of fellow patients in the same category. In the control medical ward, for example, 80 of the 86 choices made by Caucasians were for Caucasians, and 8 of the 11 choices made by Negroes were for

Negroes. Therefore, 88 out of 97 or 91% of the choices were for persons similar to self on the racial variable. The distribution of patients with these characteristics was about the same in the three wards. The general pattern of results in Table 2 suggests that as mental illness increases, similarity in social background variables plays less of a role in determining sociometric choices. On the control medical ward, four social background variables yielded significant percentages—age ($p < .03$), race ($p < .01$), rank ($p < .02$), and length of service ($p < .04$). On the open psychiatric ward, only three background variables yielded significant percentages—age ($p < .01$), race ($p < .01$), and population area of origin ($p < .01$). On the locked psychiatric ward, no social background variables were significant. The locked ward percentage is significantly lower ($p < .07$ or better) than the other two wards for age, race, and population area of origin. Rank, length of service, marital status, and education show similar trends, but these are not significant. The results suggest several additional aspects of social behavior that might be influenced by mental illness. Caudill et al. (1) also report that on an intensively studied ward there was a "... muting of outer-world distinctions on the basis of race, ethnic group, or social class ..."

Sociometric Choice and Similarity of Diagnosis

An effort was also made to evaluate sociometric choices in terms of similarity in diagnosis. When the data for the control medical ward were analyzed by amputation, fracture, and other groups it was found that 20% of the sociometric choices were within groups. The

psychiatric wards were subdivided into paranoid schizophrenic, other schizophrenic, character disorder, neurotic, and other groups. On the open psychiatric ward, 40% of the choices were within groups, a value somewhat higher than the 21% on the locked psychiatric ward. When the psychiatric wards were categorized as schizophrenic and nonschizophrenic, 62% of the choices on the open ward and 46% on the locked ward were within groups. These differences are not significant, and these diagnostic factors appear to be less important than some of the social background variables. On the other hand, Shipman's finding (5) that paranoids tend to choose one another while schizoids do not was supported. When the choices of schizophrenics from both psychiatric wards were examined it was found that 64% of the choices by paranoid schizophrenics were for other paranoid schizophrenics as opposed to all other patients. Only 33% of the nonparanoid schizophrenic choices were for nonparanoid schizophrenics. This difference is significant ($p < .05$).

Sociometric Choice and Propinquity

A somewhat different kind of variable that might be related to mental illness is propinquity, defined as the degree to which patients made sociometric choices from among those patients who slept in the same area as they. It will be recalled that each ward was divided into three sleeping areas. Table 3 shows the percentage of choices within the patient's own sleeping area on positive and negative items for the three ward groups. The percentage that would be expected if the choices were ran-

TABLE 3
PROPINQUITY AS A FACTOR IN SOCIOMETRIC CHOICES AND DEGREE OF MENTAL ILLNESS

Sociometric Measure	Degree of Mental Illness			
	Control Medical Ward	Open Psychiatric Ward	Locked Psychiatric Ward	All Wards
Percentage of positive choices within own section of ward	44	45	54	46
Percentage of negative choices within own section of ward	42	32	38	39

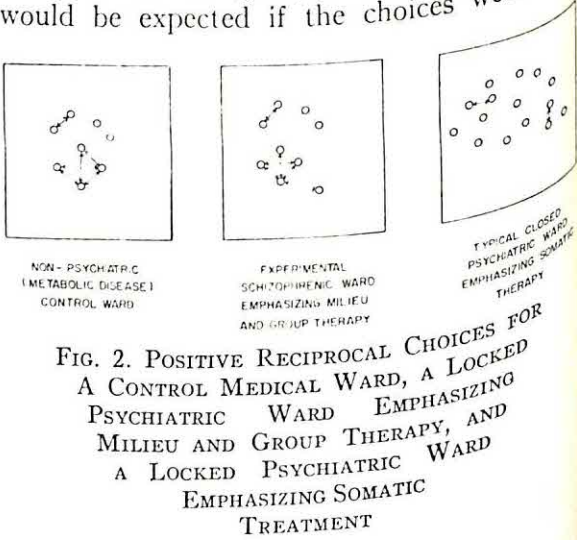


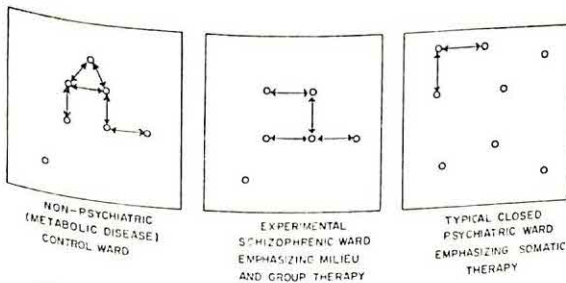
FIG. 2. POSITIVE RECIPROCAL CHOICES FOR A CONTROL MEDICAL WARD, A LOCKED PSYCHIATRIC WARD EMPHASIZING MILIEU AND GROUP THERAPY, AND A LOCKED PSYCHIATRIC WARD EMPHASIZING SOMATIC TREATMENT

domly distributed throughout the ward is $33\frac{1}{3}$. There are no significant differences between the wards on either positive or negative items. However, over all wards more positive choices are made within the patient's sleeping area than would be expected by chance ($p < .001$). On the other hand, the percentages of negative choices are what would be expected by chance. The difference for all wards between the percentage of positive and negative choices within the area approaches significance ($p < .06$). There is an insignificant trend for this difference to increase with degree of mental illness.

Milieu Therapy and Ward Organization

An initial indication of the effect of milieu therapy on social relationships can be seen in Fig. 2 where the sociograms for the three smaller wards are compared. The social organization on the nonpsychiatric control ward, as indicated by reciprocal choices on the "like-best" item, is fairly complex and involves 75% of the group. On the other hand, the somatic therapy ward shows little social organization with reciprocal choices on the "like-best" item limited to 27% of the group. The critical ward is the milieu therapy one shown in the center of Fig. 2. It can be seen that the milieu therapy ward has a complex organization involving 78% of the group and therefore resembles the nonpsychiatric control ward much more than the typical closed psychiatric ward. The difference between these percentages in the control and milieu therapy wards is not statistically

significant, while the difference between the milieu therapy and somatic therapy wards is ($p < .05$). The number of patients on these wards is too small to justify an elaborate statistical analysis. More reliance is placed on the fact that the same over-all, qualitative result was obtained when the study was replicated with the same wards after a patient turnover. This is shown in Fig. 3: the nonpsychiatric control and the milieu therapy wards show a complex social organization, while there is much less interaction and social patterning on the somatic therapy ward. It is conceivable that the lack of social organization on the somatic therapy ward is due to a direct effect of the treatment. But this is not likely. An examination of the treatment history of each individual patient on the ward failed to reveal a relationship between type or length of somatic therapy and sociometric results. For example, the three patients with reciprocal choices in Fig. 3 included a patient who had just completed 20 electroconvulsive shocks, a patient getting heavy doses of a tranquilizing drug, and a patient who had had 50 insulin coma treatments. These patients can be compared with the three most isolated patients of whom one had just completed 24 electroconvulsive shocks, one was getting heavy doses of a tranquilizing drug, and another had just had six electroconvulsive shocks. While the results suggest that milieu therapy improves group organization on the ward, other evidence is needed to evaluate the long-term effects of the milieu therapy on the patient's illness and adaptive capacities.



SUMMARY AND CONCLUSIONS

The study was designed to test the hypothesis that the greater the degree of mental illness in a patient, the more disturbed are his social relationships. A sociometric questionnaire was administered to three wards arranged along a continuum of degree of mental illness—a control medical, an open psychiatric, and a locked psychiatric ward.

A sociogram showed that the complexity of ward social organization decreased as degree of mental illness increased. Percentage of patients with one or more reciprocal choice, percentage of choices reciprocated, absolute number of names known, and the absolute number

FIG. 3. POSITIVE RECIPROCAL CHOICES FOR A CONTROL MEDICAL WARD, A LOCKED PSYCHIATRIC WARD EMPHASIZING MILIEU AND GROUP THERAPY, AND A LOCKED PSYCHIATRIC WARD EMPHASIZING SOMATIC TREATMENT (BASED ON A SECOND SOCIOMETRIC QUESTIONNAIRE AFTER A PATIENT TURNOVER)

of choices made decreased as degree of mental illness increased. The percentage of social isolates increased in the more disturbed wards.

The results suggest that as mental illness increases, sociometric choices are influenced less by similarities in social background variables such as age, race, and population area of origin. Except for paranoid schizophrenics, similarities in diagnostic variables have little effect on sociometric choices.

There were fewer reciprocals, number of choices, etc., on negative sociometric items than on positive ones for all wards. Negative reciprocals, choices, etc., decreased as degree of mental illness increased but proportionally to the decrease in positive items.

Sociograms from three additional wards showed a degree of social interaction and organization on a psychiatric ward emphasizing milieu and group therapy which was quite comparable to that of a control medical ward. In contrast, a psychiatric ward emphasizing somatic therapy showed much less interaction and organization. After a patient turnover, the study was replicated with similar results.

It was concluded that as degree of mental illness increases, there is a decrease in social organization and social relationships involving positive or negative feelings. This process appears to be reversed by milieu therapy.

REFERENCES

1. CAUDILL, W., REDLICH, F. C., GILMORE, H. R., & BRODY, E. B. Social structure and interaction processes on a psychiatric ward. *Amer. J. Orthopsychiat.*, 1952, **22**, 314-334.
2. FROMM-REICHMANN, FRIEDA. *Principles of intensive psychotherapy*. Chicago: Univer. Chicago Press, 1950.
3. McMILLAN, J. J., & SILVERBERG, J. Sociometric choice patterns in hospital ward groups with varying degrees of interpersonal disturbances. *J. abnorm. soc. Psychol.*, 1955, **50**, 168-172.
4. RIOCH, D. McK., & STANTON, A. H. Milieu therapy. *Psychiatry*, 1953, **16**, 65-72.
5. SHIPMAN, W. G. Similarity of personality in the sociometric preferences of mental patients. *J. clin. Psychol.*, 1957, **13**, 292-294.
6. SULLIVAN, H. S. *Conceptions of modern psychiatry*. Washington, D. C.: William Alanson White Psychiatric Foundation, 1947.

Received September 12, 1957.

A COMPARATIVE STUDY OF INDIVIDUAL, MAJORITY, AND GROUP JUDGMENT

DEAN C. BARNLUND

Northwestern University

THE comparative quality of decisions made by groups and by individuals working alone has been tested under a wide variety of experimental conditions. In general, these studies indicate that group judgments are superior to individual judgments on certain types of intellectual problems (2, 6). Where experiments have employed groups composed of persons of different levels of ability, however, it is not clear whether the quality of the decisions is due to the greater influence of the more capable members of the group or is a specific consequence of group thinking itself. Do groups make better decisions because the less intelligent capitulate to the more intelligent members? Or are there psychological factors inherent in group interaction which produce the higher level of performance? When each group member possesses unique information or ideas it is not unreasonable to expect that interaction will increase the total amount of information and enlarge the perspective of the group as a whole. But what happens to the level of group judgments when interaction occurs among persons who are equally informed and talented?

The present investigation is concerned with how decisions made by individuals working alone compare with the pooling of individual judgments through majority vote and with decisions reached through the process of group discussion when: (a) the membership of the group is homogeneous with respect to ability to solve the assigned problem; (b) the task is complex, couched in prejudicial terms, and involves a range of possible solutions; and (c) individuals and groups are permitted the same length of time to complete their tasks. Finally, the study seeks to determine some of the factors that account for any differences observed in individual and group performance.

METHOD

Subjects

The Ss used in this experiment were students enrolled in freshman courses in group discussion over a three-

year period at Northwestern University. The members of eight classes were used, 174 students in all. Of these, 143 were assigned to experimental groups, and the remaining 31 served as control Ss.

Procedure

At the first meeting of the classes, Form A or Form B of the "Recognition of Valid Conclusions" test was administered. Form A and Form B were alternated as the first and final measures of problem-solving ability throughout the experiment to reduce any biasing effects growing out of differences in the two forms. Each student was given a copy of the test and an answer sheet and instructed to work out his solutions to the 30 problems individually. Ss were given the 50-minute period to complete the test items.

Each member of the class was then ranked according to the total number of items he answered correctly on the first form of the test. Eight or nine weeks later, before the end of the academic quarter, experimental groups were created. Four or five groups were formed in each of the classes used in the experiment. All students who received the same or similar scores on the first test were placed together so that homogeneous groups were created. Experimental groups were then given a single answer sheet and copies of the alternate form of the test and instructed to reach a group decision on each of the 30 problems. Experimental and control Ss were again given 50 minutes to finish the test. Members of the control classes repeated the test under the original conditions, solving items on the alternate forms of the test individually. A total of 29 experimental groups participated in the experiment.

The final 10 group sessions were tape recorded in their entirety. An analysis was made of each of the 30 decisions reached by the 10 groups to isolate the specific kinds of mistakes that contributed to the majority of group errors. Following all group tests, discussions were held with the experimental Ss concerning the factors they felt influenced their performance as members of the groups.

Problem

Many investigators of group phenomena have admitted difficulty in finding or constructing suitable instruments for testing the efficiency and accuracy of group decision-making. Problems, to be realistic, should be complicated enough so they cannot be solved by intuition. They should be sufficiently difficult to test the limits of individual and group thinking. Social problems normally can be solved in a variety of ways, and test problems should contain this same feature. The difference between a right and wrong decision, however, should be clear and demonstrable. If possible, problems should be presented so as to involve the total

personality of the individual and permit his prejudices to influence his judgment as they do in a majority of everyday problem-solving experiences.

The instrument used in this experiment, the Bradley test of *Formal Validity in Problem Solving*, seemed particularly well adapted for this purpose (1). The first section of the test entitled, "Recognition of Valid Conclusions," proved long enough and sensitive enough to provide data on the experimental hypothesis. The 30 problems which make up the test consist of partially constructed arguments of varying degrees of difficulty. Two statements are given which are to be assumed to be materially true. The problem is to select the conclusion that follows most logically from the premises. The arguments cover a wide range of subjects and are phrased deliberately to complicate the decision for the reader; that is, statements involve atheists, Communists, Republicans, college professors, and other terms likely to prejudice judgment. An example of one of the problems is given below:

Some Communists are advocates of heavy taxes;
All advocates of heavy taxes are conservative Republicans;

Therefore:

- a. Some advocates of heavy taxes are not Communists
- b. Some Communists are conservative Republicans
- c. Some conservative Republicans are Communists
- d. Some Communists are advocates of heavy taxes
- e. None of these conclusions follows

The validity and reliability of the instrument has been established. The 30 problems on each of the two forms include the 19 valid moods of the syllogism along with the 11 most common fallacies. The test has been successful in discriminating among college students with different backgrounds in logic, mathematics, and problem solving. Intercorrelation of the two forms yields a raw score " r " of .85 (PE, .015) and a weighted " r " of .88 (PE, .012). Items have been carefully scaled in the final forms so that similar scores represent similar patterns of individual errors.

RESULTS

Measures of the relative effectiveness of individual, majority, and group judgments were obtained from scores made on the two forms of the "Recognition of Valid Conclusions" test.

The number of items answered correctly on the first form was used to set up homogeneous groups and to determine the level of ability represented by the average scores and "superior" scores of members of the experimental groups when working alone.¹ The relative ac-

¹ The "superior" member of a homogeneous group is something of a misnomer. Experimental groups were made up of Ss whose initial scores differed by no more than a few points. In each case the "superior" member refers simply to the individual who made the highest individual score in the group despite its homogeneous character.

curacy of problem solving under conditions of majority rule was derived from an item analysis of the individual answers of each group member. This "mathematical majority" indicated how the groups would have scored if they had pooled their opinions by secret ballot. Of the total of 829 decisions made by the experimental groups, 22 were found to be deadlocks. These occurred whenever a group of four or six Ss divided their votes equally between right and wrong answers. The results of splitting these decisions evenly and from crediting all of them to the advantage of the majority are recorded in Table 1 under "Deadlocks divided" and "Deadlocks credited." The quality of group thinking was measured by computing the mean scores of experimental groups on the second form of the test when they were required to reach consensus on each of the test problems.

The mean scores obtained under the various experimental conditions and the t values they yield are summarized in Table 1. The average scores of members of the experimental groups working alone are not significantly larger or smaller than the mean of majority scores when the 22 deadlocks are counted as correct in half of the instances and incorrect in the other half. When all deadlocks are resolved in favor of the correct decision, majority rule proves to be superior to the average performance of the individual group members. The "superior" members of the experimental groups, on the other hand, did significantly better than the majority when deadlocks were split, and as well as the majority when deadlocks were counted as correct solutions to the problems.

Group decisions were found to be clearly superior to individual decisions. As a result of discussion, experimental groups obtained mean scores that were significantly higher, at the .01 level, than "superior" members of the same groups were able to attain through individual effort. These findings also hold true when results for Form A and Form B are analyzed separately. Groups whose members scored initially near the upper limit of the test, 28 or 29 correct answers out of a possible 30, gained least from solving problems cooperatively. The largest gains were made by groups whose initial scores were low although nearly all of the experimental groups, with the exception of the highest scoring group in each class,

TABLE 1

COMPARISON OF INDIVIDUAL, MAJORITY, AND GROUP SCORES ON THE "RECOGNITION OF VALID CONCLUSIONS" TEST

				<i>t</i>
Individual Decisions		Majority Decisions		
Means of average individual scores	17.5	"Deadlocks divided"	17.9	1.73
		"Deadlocks credited"	18.3	2.72*
Means of "superior" individual scores	18.8	"Deadlocks divided"	17.9	3.23**
		"Deadlocks credited"	18.3	1.66
Individual Decisions		Group Decisions		
Means of average individual scores	17.5	Mean scores of groups	21.9	9.46**
Means of "superior" individual scores	18.8			5.77**
Majority Decisions		Group Decisions		
"Deadlocks divided"	17.9	Mean scores of groups	21.9	6.60**
"Deadlocks credited"	18.3			5.95**

* Significant at .05 level.

** Significant at .01 level.

made substantial gains as a result of group deliberation. Students in the lowest fifth of their classes as a group often rivalled the performance of the most brilliant member of the class working alone. In only two of the 29 experimental groups did students working together fail to outperform their own best member.²

When majority rule is compared with group consensus, the results show a similar large and significant advantage for group decision-making. Crediting all deadlocks from divided votes reduces the size of the group advantage over majority decisions, but its value is still highly significant.

The 31 control Ss had mean scores on the initial administration of the "Recognition of Valid Conclusions" test of 18.5. (Control Ss made an initial mean score of 18.8 on Form A and of 18.2 on Form B.) On the final test form their mean score was 18.7. (The final mean scores for control Ss were 18.9 on Form A and 18.6 on Form B.) This difference is not statistically significant and it is safe to assume that differences in mean scores of the experimental Ss were due to the experimental variables rather than differences in the test forms.

These data indicate that the members of homogeneous groups can achieve significantly better decisions by solving their problems cooperatively than they can through voting or by individual effort. Majority decisions, when all deadlocks can be successfully resolved, can produce better results than are obtained from the averaging of individual efforts. But in three out of four of the conditions observed in this experiment, majority decisions proved to

be no better than, or inferior to, the decisions of individual members of the same groups.

DISCUSSION

The results of the first phase of this experiment need to be interpreted in the light of early research on collective judgments. Whether they explained the finding on statistical or psychological grounds, Watson (10), Gordon (3), Stroop (8), and Gurnee (4) found grouped judgments superior to those of the average individual and equal to those of the superior individual working alone. This conclusion is not supported by our data. When deadlocks are resolved on the basis of statistical probabilities, majority decisions are found to be no better than those of the average member of homogeneous groups.

The explanation for the difference in results seems to lie partly in the character of the tasks and partly in the methods of grouping data. Some of the problems used by these investigators involve what may be called additive activities. Whenever individual efforts are additive or cumulative, the larger the group the greater should be the advantage from combining individual data. Testing the accuracy of conclusions drawn from given arguments is not the same kind of problem. One answer simply cannot be added to another. A second explanation for the difference is found in the manner of grouping individual decisions. The pooling of data in previous studies combined the heterogeneous opinions of 10 to a 100 individuals. In averaging data the greater the number and range of scores, the larger the gain from cancelling out individual errors. In this case, only four to six opinions from individuals of comparable ability determined the

² In both of these cases the groups contained individuals who received almost perfect initial scores.

decision. Majority rule may prove a convenient political device for averaging individual preferences; but our results suggest that in small, homogeneous groups or committees, majority rule, when it precludes discussion or debate, is likely to be less effective than the personal judgment of superior members of the group.

After discussion, however, experimental groups produced decisions that were far superior to those of members working alone or through majority rule. Moreover, group decisions on the test problems were reached within the same period of time allotted to individuals.

Several hypotheses are offered in the literature for the high quality of group judgments. Watson found group decisions superior because of the influence of the ablest member.

In measuring the output of a group, either when working along cooperative group-thinking lines or when the project permits the simple compilation of individual efforts, it matters little about the ability of the poorest or even average member of the group. The results seem to show primarily what the few ablest in the group have produced (10, pp. 333-334).

This hypothesis, though generally tenable, seems inadequate to explain the results of this experiment. Groups were made up of students whose initial performance indicated a common aptitude for selecting logical conclusions from given arguments. The grading of items on the "Recognition of Valid Conclusions" test is such that persons who get similar scores are likely to possess not only the same level of ability but similar habits of thinking.

Another theory, suggested by Gurnee (4) and Thorndike (9), is that the better performance of the group is due to the social influence of the more confident group members who are more often right than wrong. It is difficult to see how this factor could have played a large part in the results. It would seem likely that students with similar patterns of right and wrong answers would share somewhat similar patterns of confidence about their answers. If so, this factor can be minimized.

It is necessary to go beyond these hypotheses to explain how correct solutions were reached by groups whose members made similar or identical errors when working alone. The diagnostic discussions and the analysis of re-

corded group sessions furnish additional clues to the psychological factors affecting the high level of group performance.

Membership in the experimental groups produced a higher level of interest in the successful completion of the task. Ss concentrated more intently on the assigned problems after being appointed to a group than they did when solving the problems individually. Group members found themselves more and more deeply involved as they proposed, and were forced to defend, their ideas. Participants identified with their own groups to such a degree that when some members became fatigued, others urged them to continue working.

Membership in the experimental groups had an inhibiting as well as facilitating effect. Knowledge that one's opinions were to be shared publicly made group members more cautious and deliberate in their own thinking. The necessity of explaining a conclusion forced many students to be more self-critical. Errors that might have been committed privately were checked before they were communicated to others.

Groups had greater critical resources than did individuals working alone. In spite of the uniform level of ability, group members saw different issues and a larger number of issues than a single person did working alone. A greater number of viewpoints increased the group's chances of selecting a valid one. Even the poorest members contributed significantly to the quality of the group product. Remarks that went no deeper than "I don't understand" or "That's absurd" often saved the group from error by forcing others to justify their opinions, and in so doing disprove their own conclusions.

A more objective view of the problem resulted from competition between the private prejudices of group members. The test arguments were stated in loaded terms designed to make the choices between conclusions as difficult as possible. Each individual, however, brought a different set of values to his group. When arguments were stated so they appealed to persons of one persuasion, those in opposition were anxious to detect their error. In this way, liberals counteracted conservatives, Republicans offset Democrats, and "independents" guarded against critical lapses on the part of fraternity members. Groups were forced to

become more objective, and this, of course, increased their chances of drawing valid conclusions. The significance of this one factor alone would be hard to overestimate.

Discussion of the test items also prevented other incidental mistakes from occurring. Some groups had to check their instructions several times because members had different interpretations of them. Discussion often led to a clarification of terms used in the test, and, where logical fallacies spring from ambiguous terms, this may account for some of the gains. A number of groups formulated general principles as they went along to help them avoid repeating errors in later problems.

What, then, prevented experimental groups from attaining even higher scores than they did? Analysis of the transcripts revealed two factors that together accounted for a majority of the group errors. The first was that group members agreed immediately and unanimously upon the wrong answer to a problem. Further study of the issue was then considered unnecessary and wasteful. This is the same factor that Jenness, following F. H. Allport, refers to as the "impression of universality" (5). Agreement becomes the criterion of correctness. Maier (7) suggests that provoking arguments under these circumstances leads to better judgments. The virtue of disagreement and the possible function of a "No-Man" in group deliberations, needs further testing.

The second factor was that groups, when they reached a deadlock, were unable to use their differences of opinion for their own advantage. When conflicts became intense they were resolved by surrender of the less aggressive members or by compromising on a third solution which was almost always incorrect but served to protect the egos of the parties to the controversy. Apparently disagreement stimulates thought up to a point; beyond that point, groups may lack the patience and skill to exploit it.

Discussion, as a preliminary to group decisions, causes groups to examine a problem more thoroughly and to consider a wider number of solutions. It encourages individuals to think more carefully and in sharing opinions to expose the logic of their position to the inspection of others. Membership in a group produces a sense of responsibility which in-

tensifies and sustains effort. The biasing effect of private prejudice may be counteracted leading to a more objective view of the issues. The data of this study indicate that the answer to the question of whether group opinion is better than individual opinion because of the influence of the superior person or because of the discussion process itself is that discussion inherently contains psychological pressures and motivations which, if not abused, tend to produce superior judgments on complex intellectual problems. Individual decisions and collective judgments lack the additional ingredient supplied by interaction which permits a group to outperform its own members.

SUMMARY

The performance of individuals working alone, under majority rule, and as members of discussion groups were compared on a complex intellectual task. Individual judgment was measured by administering a test of ability to draw logical conclusions from given arguments. Individuals receiving similar scores were assigned to the same experimental groups so that the factor of distributed ability would be reduced to a minimum. The votes of members of the homogeneous groups were mathematically tallied to determine the results under conditions of majority rule. A second form of the test was completed as a group undertaking and the scores compared with individual and majority scores. The results indicated that:

1. Majority decisions, when deadlocks are evenly divided between right and wrong answers, are not significantly different from those made by the average individual and are inferior to those of the best member of the group working alone.

2. Group decisions, reached through cooperative deliberation, are significantly superior to decisions made by individual members working alone and to majority rule.

The superiority of group judgments was found not to be a reflection of the wisdom of the superior member of the group but a result of psychological factors inherent in discussion. Participation in a group led to more serious concentration on the task and to more enthusiastic individual effort. Group discussion was found to stimulate more careful thinking, to lead to a consideration of a wider range of

ideas, and to provoke more objective and critical testing of conclusions.

REFERENCES

1. BRADLEY, E. E. Formal validity in problem solving. Unpublished doctoral dissertation, Northwestern Univer., 1950.
2. DASHIELL, J. F. Experimental studies of the influence of social situations on the behavior of human adults. In C. Murchison (Ed.), *Handbook of social psychology*. Worcester: Clark Univer. Press, 1935. Pp. 1097-1158.
3. GORDON, K. H. Group judgments in the field of lifted weights. *J. exp. Psychol.*, 1924, **7**, 398-400.
4. GURNEE, H. A comparison of collective and individual judgments of fact. *J. exp. Psychol.*, 1937, **21**, 106-112.
5. JENNESS, A. The role of discussion in changing opinion regarding a matter of fact. *J. abnorm. soc. Psychol.*, 1932, **27**, 279-296.
6. KELLEY, H. H., & THIBAUT, J. W. Experimental studies of group problem solving and process. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954. Pp. 735-785.
7. MAIER, N. R. F., & SOLEM, A. R. The contribution of a discussion leader to the quality of group thinking. *Hum. Relat.*, 1952, **5**, 277-288.
8. STROOP, J. R. Is the judgment of the group better than that of the average member of the group? *J. exp. Psychol.*, 1932, **15**, 550-562.
9. THORNDIKE, R. L. The effect of discussion upon the correctness of group decisions, when the factor of majority influence is allowed for. *J. soc. Psychol.*, 1938, **9**, 343-362.
10. WATSON, G. B. Do groups think more efficiently than individuals? *J. abnorm. soc. Psychol.*, 1928, **23**, 328-336.

Received September 12, 1957.

THE EFFECT OF ATTITUDE AND EXPERIENCE ON JUDGMENTS OF CONTROVERSIAL STATEMENTS¹

MARSHALL H. SEGALL

Columbia University

IN SO FAR as opinion formation and change involve subjective judgments, several aspects of adaptation-level² theory (4) should prove to be of value for the development of a psychology of opinion. Among the most promising conceptions incorporated in Helson's unification of judgment principles are those of the implicit relativity of absolute judgment, a neutral reference point in terms of which incoming stimuli are evaluated, and the level of adaptation as a summary of previous relevant stimulation.

There is considerable agreement that a more solid basis is needed for a theory of attitudes, social norms, frames of reference, and the like, and since A-L theory has been offered as such (5), it seems well to assess the relevance of the theory to problems in this area. This study is focused on the effects of prior experience on extremity evaluation of opinion statements. According to the theory, absolute extremity judgments, like those required in the construction of a Thurstone questionnaire, are relative to each judge's A-L. It follows that such so-called absolute judgments are in fact relative to the judge's attitude, at least to the degree that his attitude reflects his experience with material similar to that being judged. In this regard, it is interesting to note that a relationship between attitude and judgment of controversial material has recently been found by Hovland and Sherif (8). Thurstone's assumption of independence of attitude and judgment in this setting (12), though seemingly upheld by the work of several investigators (e.g., 3, 7, 10), has now been seriously

challenged by the Hovland and Sherif demonstration of displacement of judgment away from the judge's own position. Hovland and Sherif note that this finding is virtually demanded by the well-documented principles of judgment which emerged over the years from the psychophysics laboratories. These principles are the foundation of A-L theory.

In the present study, an effort was made to replicate the Hovland and Sherif finding with similar material and, in addition, to test the utility of postulating an underlying A-L mechanism. To this end, in addition to employing judges known to hold different positions on the topic of the items to be judged, the nature of the experience with the items during the judgment session was systematically manipulated. To demonstrate the strict relevance of A-L processes, it seems necessary first to demonstrate that experimentally manipulated experience can lead to displacement of judgment.

According to Helson's theory, *background* and *residuals* interact to establish the A-L which serves as the reference point determining the response to each focal stimulus. As each stimulus is judged, it then becomes part of the background, thereby modifying the A-L as effective in subsequent judgments. As Helson's theory is here interpreted, variations in judges' attitudes represent variations in residuals, and manipulation of the order of occurrence of items to be judged is an empirical translation of variation in background.³

The study was so designed that the relative influences of residual and background factors

¹ This report is based on a dissertation submitted to the Graduate School, Northwestern University, June, 1957. The writer wishes to express his gratitude to Donald T. Campbell for his guidance and encouragement during all phases of this investigation, and to Donald J. Lewis, Carl P. Duncan, William A. Hunt, and Brenden Maher for their many contributions. Rudolph Schulz, Donald Butler, and Norman Miller also offered several helpful suggestions during the course of the study.

² Hereafter abbreviated A-L.

³ These identifications of residual and background classes of stimulation with judge's attitude and item order seem consonant with Helson's intentions. "Although the distinction between stimulus as such and background stimulus is not always easy to draw, when an individual is instructed to respond to specific stimulation during or soon after other stimuli have acted, the latter may be regarded as background stimuli with respect to the focal stimuli. Beliefs, attitudes, traits, and cultural determinants which individuals bring into any concrete situation may be regarded as residuals in affecting specific responses" (6, p. 314).

in each of two judgment tasks could be determined. In one task, absolute judgments of the extremity of statements about college fraternities were required (this is essentially the Thurstone task); in the other, judges were required to state the extremity of statements in comparison with their own positions. The latter task was introduced in order to detect the presence of A-L phenomena when relative judgments are explicitly called for.

In addition, the design of the study permits detection of a context effect. A-L theory predicts that neutral items when judged in context with *anti* items will be judged as closer to the *pro* end of the attitude dimension than when judged in a *pro* items context. In the latter instance, judgments should be displaced toward the *anti* end of the dimension.

Finally, this study provides data relevant to a prediction implied by A-L theory of a measurement bias inherent in the Thurstone approach. If the range of positions spanned by the questionnaire items differs sufficiently from the respondent's own position, so that his A-L is shifted, his responses to those items would be modified, thereby producing a change in attitude score. The second task employed here was arranged so that a change of this nature, were it to occur, could be detected. Whether this kind of change in score reflects a change in attitude is a moot question; a change in score due only to questionnaire form and content is of sufficient methodological concern to warrant study.

METHOD

A factorial design was employed, incorporating three groups of judges known to differ in attitude toward college fraternities, three orders of presentation of statements about fraternities, two judgment tasks, and an additional variation in manner of item presentation involving the inclusion or noninclusion of "transition" items.

Of the total 180 Ss, 60 were classified "pro-fraternity," 60 "moderate," and 60 "anti-fraternity." In one order condition, 24 pro-fraternity statements preceded 24 anti-fraternity statements; in a second condition that order was reversed. In a third, or control, condition, both kinds of statements were intermingled. In all three conditions, 16 neutral items were interspersed among the pro and anti statements. In addition, for one half of the Ss a few additional items were included after the thirty-second item to provide a smooth transition phase between halves of the total item complement. Finally, the task variable was introduced by instructing one half of the Ss to make absolute judgments,

while the other half were told to indicate the statements that best represented their own positions and to judge the extremity of all other statements in terms of those positions. These variations, to be discussed in detail in the succeeding sections, resulted in a 36-cell experimental design.

Subjects

Ss were drawn from a pool of 227 introductory psychology students who had completed a 120-item multi-topic attitude questionnaire several weeks before the experiment proper began. The questionnaire was administered in class by the regular instructor and was in no way identified as part of this experiment. That the questionnaire was never mentioned by the Ss indicates that they did not link the two events. On the basis of their responses to seven imbedded items dealing with fraternities, 70, 89, and 68 respondents respectively were classified pro, moderate, and anti. One month after the questionnaire administration, daily circulation of volunteer appointment sheets was begun. This method of recruiting Ss is standard, so that volunteering for this study was, from the students' point of view, not unusual. However, unknown to them, *E* possessed classification data, and as Ss volunteered they were assigned to the appropriate attitude group.

Within each attitude group, Ss were assigned in rotation to one of 12 subgroups until each of the 36 cells in the design matrix contained five Ss. Efforts were made to insure that all cells were filled at the same rate. Occasionally it was necessary to telephone potential Ss in a given attitude classification, but selection biases were minimized by obtaining in this manner an approximately equal number from each classification.

Materials

One hundred and fifty statements collected from various sources were categorized by ten psychology graduate students along a seven-category continuum ranging from extremely anti (Category 1) to extremely pro (Category 7). Sixteen neutral items (Category 4) and ten items each at the other six points were retained for use in the sorting tasks. (Henceforth, the category values assigned by the graduate students will be referred to as "true values.") Each of the 76 statements was typed on 3 x 5 cards to facilitate the variation in order of presentation. Each S was provided with a set of at least 64 cards arranged in a predetermined order. Thus, for those Ss in the *con-pro* condition, the first 32 cards contained statements that were *con* or neutral (true values 1 through 4) and the last 32, *pro* or neutral (4 through 7). The reverse was true for the *pro-con* condition. In the remaining order condition, the whole range of values was represented in both the first and last 32 items.

A seven-compartment card receptacle with narrow slits along the top was employed to make it impossible for S to recover a card once sorted, or to see how many cards he had assigned to each compartment. The compartments were numbered from 1 to 7 and labeled appropriately for the particular sorting task being performed. The numbers were fixed, but the labels in-

cating the meaning of each compartment were posted just prior to each *S*'s appearance at the laboratory.

Tasks

Each *S* was instructed to perform one of the two card-sorting tasks. Random selection determined the pairing of *S*s and tasks, and both tasks were run concurrently to avoid confounding the task variable with extra-experimental events. Extremity judgments were required in both tasks. In the absolute judgment task, Category 4 was defined as objectively neutral, while in the relative judgment task, that category was reserved for statements which *S* saw as expressing his own position. In the former task, the six other categories were labeled: [1] *extremely anti-fraternity*, [2] *moderately anti-fraternity*, [3] *slightly anti-fraternity*, etc., through to [7] *extremely pro-fraternity*. In the latter task, those categories were labeled as follows: [1] *very much too unfavorable toward fraternities to represent my feelings*, etc., through to [7] *very much too favorable . . . to represent my feelings*. It should be noted that the latter task is similar to that performed by respondents to an attitude questionnaire, with the additional requirement that rejected items be arrayed along the extremity continuum.

Procedure

Order of presentation. A basic packet of 64 cards in the *pro-con* order was arranged as follows: The first 32 cards consisted of eight items each of true values 4, 5, 6, and 7, and the last 32 were made up of eight items each of values 1, 2, 3, and 4. Within each group of 32, items were arranged so that each true value was represented once in every successive four items. Within each set of four, the values represented occurred in any one of 24 orders, the permutations of four values taken four at a time. These precautions were taken to insure that the only systematic true value variation was between sets of 32, with only random variation within. An additional control, for differential impact of individual statements, was employed by selecting arbitrarily for each *S* 48 *pro* and *con* statements from the available 60. The ten available items in each of the six *pro* and *con* categories were shuffled, and the eight topmost cards were used in the order determined by the shuffle. The two remaining cards were omitted from the basic packet of the *S* in question. Thus, all *S*s in the *pro-con* condition received 32 cards of values 4 through 7 followed by 32 cards of values 1 through 4, but each *S* in that group actually received a unique order of statements.

The general format prevailing in the *con-pro* condition was the direct reverse of that used in the *pro-con* condition, with the same controls in effect. A similarly unique order of items was achieved for *S*s in the control condition, but here, all true values were represented throughout each packet. Thus, every successive arrangement of eight items included one statement each of values 1, 2, 3, 5, 6, and 7 and two statements of value 4. It should be clear that all *S*s in each order condition ultimately were exposed to a single packet of at least 64 items drawn from the seven stimulus categories and that no two *S*s received the same statements in identical orders.

TABLE 1

SUMMARY OF ANALYSIS OF VARIANCE OF MEAN ABSOLUTE JUDGMENTS

Source	df	MS	F	p
Judge's attitude	2	.065	4.33	< .05
Item order	2	.530	35.33	< .001
Transition	1	.003		
Interaction: att. \times order	4	.018		
Interaction: att. \times trans.	2	.005		
Interaction: order \times trans.	2	.030		
Interact.: att. \times order \times trans.	4	.042	2.80	< .05
Replications	72	.015		

Transition. Ninety *S*s, half of the total complement in each order condition, received six statements in addition to the crucial 64 received by all *S*s. These items, two each of values 3 and 5 and one each of values 2 and 6, were inserted between the first and last 32 essential items. These transition items were drawn from the cards omitted as a result of the shuffles described above, and were themselves shuffled to determine the six to be inserted. They were included for only half the *S*s to permit an evaluation of the effectiveness of transition items in masking the change in item values occurring between *pro* and *con* halves in the two experimental conditions. It might be expected that a transition phase, by minimizing *S*'s awareness of the value difference across halves, would provide more favorable conditions for the operation of context-contrast effects. In a recent study of context effects on a judgment of musical tones (2), a number of *S*s responded to a context shift with an apparent assimilation error, an effect that might be explained by *S*'s awareness of *E*'s purpose.

Neutral items. In order to determine whether shifts in judgments of common stimuli could be effected by context changes, neutral items were included in both halves of all packets. In addition, for 90 *S*s the neutral items in each half were identical, while for the other 90 *S*s a different set of eight occurred in each half. This additional minor variation was introduced in order to determine whether identical items would be as susceptible to context shifts as similar, but not identical, statements.

Data collection. Following the steps outlined above, a packet of cards was prepared for each *S* just prior to his arrival. *S*s were run individually. Instructions⁴ were read aloud by *E*, then *S* was given the packet to sort. Following his departure, results of his sort were recorded, the container was prepared for the next *S*, and a new packet was assembled. This sequence required about 20 minutes per *S*. The data were collected during six weeks in the winter of 1956.

⁴ Instructions are reproduced in their entirety elsewhere (11). The statements employed are also available there.

RESULTS

Results obtained for each task will be considered separately.

Absolute Judgment Task

Mean category score. Hovland and Sherif (8) found that pro-Negro Ss, when judging items pertaining to Negroes, tended to place an overabundance of items in the *extremely anti* category. A tendency on the part of anti-Negro Ss to concentrate items at the *pro* end of the continuum was also noted. This relation between attitude and judgment may be viewed as a manifestation of response displacement away from A-L. In the present study the same mechanism would be expected. It would be predicted that pro and anti judges would displace their judgments in opposite directions, with a higher mean response for the anti judges (due to relatively low residuals), a lower mean for pro judges (relatively high residuals), and an intermediate mean for moderate judges. In addition, as a result of the establishment of different backgrounds via the order manipulation, differences should obtain along the order dimension. Judges in the *pro-con* condition should displace their judgments in the anti direction, while judges in the *con-pro* condition should do the opposite, reflecting A-Ls on the pro and anti side of neutral. To test these predictions, a $3 \times 3 \times 2$ analysis of mean category scores was performed. Both major variables, judge's attitude and item order, resulted in significant *F* ratios, when tested against the replication mean square. It should be noted in Table 1, however, that the level of significance for the order effect is higher than that for the

TABLE 2

SUMMARY OF *t* TESTS, PRO-CON vs. CON-PRO

	Mean responses to stimuli from each true value category, absolute judgment					
	1	2	3	5	6	7
\bar{X}_{p-o}	1.11	1.61	2.54	5.13	6.07	6.54
\bar{X}_{c-p}	1.29	1.96	3.03	5.37	6.26	6.77
$s\bar{X}-\bar{X}$.066	.117	.084	.070	.101	.126
<i>t</i> *	2.73	2.99	5.83	3.43	1.88	1.83
<i>p</i>	<.005	<.005	<.005	<.005	<.05	<.05

* *df* = 58.

effect of judge's attitude, leading to a relative lack of confidence in the reliability of the attitude effect.⁵ Nonetheless, differences along both major dimensions were in the predicted direction.

Effects on stimuli of each true value. Figure 1 shows the effect of item order on mean judgments of each class of items separately. Ss exposed first to pro statements judged anti statements as *more* anti than did Ss judging these statements at the outset. Conversely, pro statements were judged as *more* pro by Ss with prior exposure to anti statements than by Ss who judged pro statements first. Ss in the control group made intermediate judgments. Differences between mean judgments of every nonneutral class of items were shown to be significant by *t* tests summarized in Table 2. (An analysis of responses to neutral items is discussed separately below.)

As shown in Fig. 2, pro, moderate, and anti judges made virtually identical judgments to stimuli throughout the range. No significant differences obtained for any stimulus class.

Neutral items. A 3×3 analysis of variance of mean judgments of neutral items occurring in the first half of each packet revealed that the context in which those items were judged was, as predicted, a significant determinant. Judge's attitude was not. This analysis is summarized in Table 3. However, the context prevailing in the second half of the judgment session did not affect judgment of neutral items as it did during the first half. The mean judgment of the *pro-con* group was the same during both halves; the mean judgment of the *con-pro*

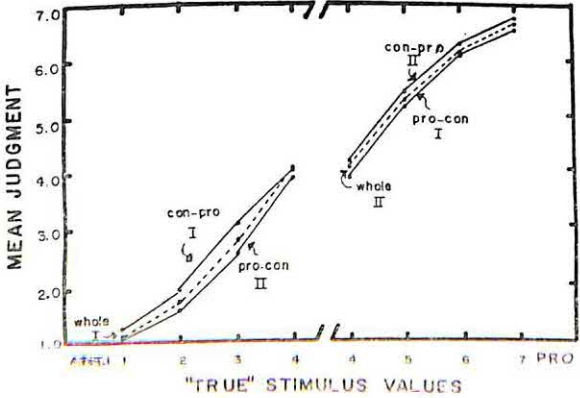


FIG. 1. MEAN ABSOLUTE JUDGMENT OF EACH ITEM CLASS BY Ss IN EACH ORDER CONDITION

⁵ Tests based upon the components of variance model (9, p. 332) lead to the conclusion that no significant difference was produced by the attitude variable.

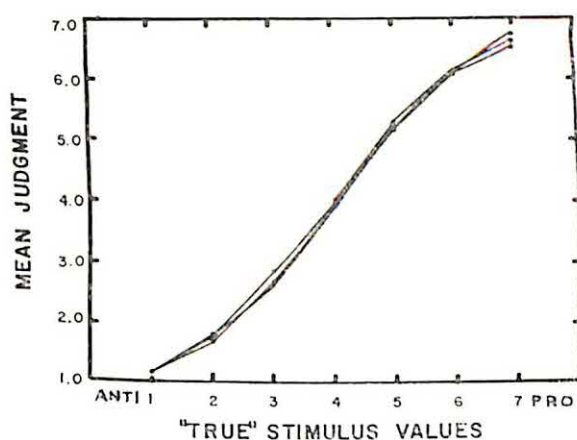


FIG. 2. MEAN ABSOLUTE JUDGMENT OF EACH ITEM CLASS BY Ss IN EACH ATTITUDE CLASSIFICATION

group shifted in the pro direction in face of the pro context, just contrary to the A-L prediction. Thus, the intrasession shifts predicted for neutral items judgment did not occur.

Relative Judgment Task

Mean category score. Analysis of data collected in this task proceeded as above. Evidence of a general displacement of judgment was first sought. An analysis of variance of mean category score (see Table 4) showed that in this task, item order was *not* related to the Ss' treatment of items. On the other hand, and again in contrast with the effects noted for the absolute judgment, the judges' attitudes were to a large extent responsible for the category assignments. However, this fact is less than startling, for in this task each S was explicitly judging the items in relation to his own position. The relationship between attitude and judgment here indicates simply that instructions were followed. A more interesting finding is the absence of an order effect, particularly since one was found for absolute judgment. When the mean responses to each class of stimuli were arranged separately for each order condition and for each attitude group, *t* tests indicated significant differences only as a function of attitude. This relationship between judge's attitude and relative judgment holds as well for neutral items.

Neutral items. In this task, the data showing the effect of order on neutral items present a complex picture. Mean judgments during the first half under both experimental conditions

TABLE 3
SUMMARY OF ANALYSIS OF VARIANCE OF NEUTRAL ITEM ABSOLUTE JUDGMENTS DURING PHASE 1

Source	df	MS	F	p
Judge's attitude	2	.07	5.05	< .01
Item order	2	.46		
Interaction: attitude \times order	4	.03		
Replications	81	.09		

TABLE 4
SUMMARY OF ANALYSIS OF VARIANCE OF MEAN RELATIVE JUDGMENTS

Source	df	MS	F	p
Judge's attitude	2	3.40	37.44	< .001
Item order	2	.02		
Transition	1	.02		
Interaction: att. \times order	4	.16	1.79	
Interaction: att. \times trans.	2	.15	1.65	
Interaction: order \times trans.	2	.19	2.09	
Interaction: att. \times order \times trans.	4	.28	3.06	< .05
Replications	72	.09		

TABLE 5
SUMMARY OF ANALYSIS OF VARIANCE OF ATTITUDE SCORES

Source	df	MS	F	p
Judge's attitude	2	6.260	37.04	< .01
Item order	2	0.345	2.04	n.s.
Interaction: att. \times order	4	0.125		
Replications	81	0.169		

were not significantly different. Thus, the predicted context effects did not occur. During the second half, under con and pro contexts, respectively, mean neutral item judgments were 3.35 and 3.58, yielding a difference that approaches significance, but this difference is in the "wrong" direction. This latter fact suggests a context-assimilation effect during the second phase of judgment, but interpretation of these data is hindered by the fact that the control group means in both phases were higher than those of either experimental groups. The relationship between context and neutral item judgment in this task remains undefined.

Effect of item order on measured attitude. It was suggested that item order could alter

attitude questionnaire scores as a manifestation of a shifting A-L. This suggestion was tested by an analysis of variance of attitude scores (mean value of items assigned to Category 4, "my own attitude"). The analysis, summarized in Table 5, indicates no significant difference in scores as a function of item order.

On the other hand, there were differences in score along the judge's attitude dimension, a finding which indicates that the attitude classification procedure applied in this study was reasonably valid. However, it is important to note the small range of average scores of the three attitude groups; the means for the pro, moderate, and anti Ss respectively were 4.89, 4.52, and 3.89.

DISCUSSION

Absolute Judgment

The principal finding of this study is that the prior presentation of attitude statements covering only part of the total stimulus dimension biased the absolute judgments of all statements. As suggested in the introduction to this paper, an effect of this nature is understandable in terms of the establishment of differential A-Ls resulting from different backgrounds. In addition, A-L theory was interpreted to predict that different residual factors, as reflected in the judge's attitude, would also bias the absolute judgment, but the data here reported indicate that attitude has little effect. It would be tempting, then, to conclude that residual factors play only a minor role in absolute judgment, except that the judgment literature contains several demonstrations to the contrary. In addition to the Hovland and Sherif paper (8), Tresselt's study (13) of weight judgments by athletic weightlifters and professional watchmakers is interpretable as evidence that past experience with weights outside the laboratory influences judgment in the laboratory. To account for the discrepancy between the present findings and those reported by other investigators, it may be well to consider the possibility that the residual factor was not adequately varied in the present case.

For example, the judges in the Hovland and Sherif study included conservative Southerners and liberal Northerners; in the present study, judges were all students at a university that

supports the fraternity system. The range of relevant experiences among the latter was, in all likelihood, far less broad than that among the Hovland and Sherif judges. Indeed, it was shown above that the range of attitudes of the judges used here was quite small, and the failure to find judgment differences related to this variable is reasonably attributable to that narrow range.

Turning now to the neutral items absolute judgment, it should be stressed that the finding of a context-contrast effect during the first half of the session is in accord both with A-L theory and the scores of psychophysical studies that it synthesizes. Less easily incorporated into this framework, however, is the fact that similar context effects did not occur during the second half of the session. At least three interpretations of the data seem reasonable. (a) Once having assigned eight neutral items to categories during Phase 1, Ss assigned similar (and in some cases identical) neutral items to the same categories during Phase 2. This hypothesis is labeled *commitment*. (b) S reacted to the contrasting context operative during Phase 2 by assimilating his judgments to that context. This is a *context-assimilation* hypothesis. (c) The experience during Phase 1 alone determined the A-L, so that judgments of neutral items during both phases were relative to that A-L and unaffected by the stimuli occurring in Phase 2. This is a *primacy* hypothesis.

Some evidence for the commitment hypothesis may be drawn from the fact that those Ss who judged two identical sets of neutral items did indeed provide almost identical judgments during both phases. On the other hand, the data for Ss judging different sets of neutral items are more consonant with an assimilation hypothesis. These Ss in both experimental groups differed as predicted in Phase 1, but then their judgments diverged rather than converged in Phase 2 when the contexts were reversed. Examination of individual protocols revealed that about half of these Ss made judgments that could be described as assimilative. In two studies by Campbell et al. (1, 2), in which the prevailing phenomenon was a contrast effect in response to a shift in context, a significant number of judges were assimilators. It will be recalled that an attempt was

made here to seek an explanation for assimilation in terms of *S*'s awareness of *E*'s intent in shifting the context. By partially masking that shift for some *S*s it was felt that assimilation would be partially eliminated. If it did occur, it would be expected more frequently among *S*s for whom the shift was abrupt and probably obvious. This was not the case. Assimilation could not be linked with any known stimulus or subject variable, thus leaving the matter of assimilation unexplained.

The primacy hypothesis admits that the data may be described in terms of assimilation, but suggests that they be "explained" in terms of contrast. This hypothesis requires that A-L theory be modified to allow temporal weighting of stimuli, with those occurring earlier in a contiguous series contributing more heavily to the A-L. A within-session primacy hypothesis coupled with a between-sessions recency notion would account for all the facts noted in the absolute judgment part of this study. These paired notions first lead to the prediction that an A-L built up over a long series of experiences in the relatively distant past would be easily modified at the outset of a new and distinguishable judgment experience. Aside from the support provided by the present demonstration that judge's attitude played a minor role in influencing judgment, the fact that weightlifters and watchmakers in Tresselt's study converged upon a common judgment scale provides additional evidence. Finally, the assumption of a within-session primacy effect accounts for the apparent assimilation found here in Phase 2, but since the hypothesis was designed to account for these data, it will require future independent confirmation.

If two studies by Campbell et al. (1, 2) are jointly considered, there is further reason for correcting that part of A-L theory which posits an equal contribution from all stimuli. In one study (2), judgments of musical pitch revealed a recency effect; in the other (1), in which judgments of degree of disturbance indicated by schizophrenics' definitions were studied, a primacy effect was found. This apparent contradiction would lead to the hypothesis that the relative likelihood of finding primacy or recency effects depends upon the nature of the material being judged. Judgment of verbal

materials, which presumably involves central processes more than does judgment of musical pitch, may be subject to recency effects. At any rate, an assumption of equal weights appears to be an oversimplification.

Relative Judgment

The data emerging from this part of the study require little amplification. As a result of instructions to judge the statements in relation to one's own attitude, judgments were free from the context influences found for absolute judgment. In a real sense, the reference point provided by "my own attitude" replaced the A-L. Helson (5) has also reported that when judgments are in terms of a comparison between variables and a standard, the A-L is largely determined by the standard.

SUMMARY AND CONCLUSIONS

An interpretation of adaptation-level theory suggests that judgment of controversial statements is determined by the judge's relevant stimulus history. Among the implications of this suggestion are (a) a dependence of absolute extremity judgments upon the attitude of the judge, and (b) a potential distortion of attitude scores due to the item order on a questionnaire.

The object of this study was to determine some of the relationships among attitude, experience, and judgment of attitude statements. Absolute judgments of statements about fraternities were made by 90 college students. Another 90 were instructed to judge the statements in relation to their own attitudes toward fraternities. Judges were classified in terms of their attitudes, measured about one month prior to the experiment proper, and were assigned so that all classifications were equally represented in both judgment tasks. Experience within each task was varied by presenting two blocks of items, each covering half the stimulus range, in one or the other order. A factorial design incorporating judge's attitude and item order was employed.

In the absolute judgment task, the order variation resulted in significant contrast effects, while judge's attitude had little influence. In the relative judgment task, responses were related to attitude, but not to item order. Attitude scores were not affected by variation in

item order. In general, findings were consistent with adaptation-level theory, provided that experiences are weighted differentially according to (a) their degree of remoteness in time, and (b) the nature of the judgment task.

REFERENCES

1. CAMPBELL, D. T., HUNT, W. A., & LEWIS, NAN A. The effects of assimilation and contrast in judgments of clinical materials. *Amer. J. Psychol.*, 1957, **70**, 347-360.
2. CAMPBELL, D. T., LEWIS, NAN A., & HUNT, W. A. Context effects found with a judgmental language that is absolute, extensive, and extra-experimentally anchored. Unpublished manuscript, Northwestern Univer., 1957.
3. FERGUSON, L. W. The influence of individual attitudes on construction of an attitude scale. *J. soc. Psychol.*, 1935, **6**, 115-117.
4. HELSON, H. Adaptation-level as frame of reference for prediction of psycho-physical data. *Amer. J. Psychol.*, 1947, **60**, 1-29.
5. HELSON, H. Adaptation-level as a basis for a quantitative theory of frames of reference. *Psychol. Rev.*, 1948, **55**, 297-313.
6. HELSON, H., BLAKE, R. R., MOUTON, JANE S., & OLMSTEAD, J. A. Attitudes as adjustments to stimulus, background, and residual factors. *J. abnorm. soc. Psychol.*, 1956, **52**, 314-322.
7. HINKLEY, E. D. The influence of individual opinion on construction of an attitude scale. *J. soc. Psychol.*, 1932, **3**, 283-296.
8. HOVLAND, C. I., & SHERIF, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47**, 822-832.
9. MCNEMAR, Q. *Psychological statistics*. (2nd ed.) New York: Wiley, 1955. P. 332.
10. PINTNER, R., & FORLANO, G. The influence of attitude on the scaling of attitude items. *J. soc. Psychol.*, 1937, **8**, 39-45.
11. SEGALL, M. H. Attitude and adaptation-level: The effect of experience on extremity judgments and on expression of opinion. Unpublished doctoral dissertation, Northwestern Univer., 1957.
12. THURSTONE, L. L. The measurement of social attitudes. *J. abnorm. soc. Psychol.*, 1931, **26**, 249-269.
13. TRESSELT, M. E. The effect of the experience of contrasted groups upon the formation of a new scale of judgment. *J. soc. Psychol.*, 1948, **27**, 209-216.

Received September 12, 1957.

INFLUENCE OF FOUR TYPES OF DATA ON DIAGNOSTIC CONCEPTUALIZATION IN PSYCHOLOGICAL TESTING

WILLIAM F. SOSKIN^{1,2}

Joint Commission on Mental Illness and Health, Cambridge, Massachusetts

RECENT reviews of the literature on diagnostic testing reflect a shift in the past decade from the search for more subtle and more "revealing" instruments to more intensive study of the behavior of appraisers in the diagnostic situation.

The present research is one of a series investigating the influence of different types of data on the conceptualizing process in diagnostic situations. Its objective is to study the differential influence of four types of data on the judgments of individuals as reflected in their responses on a criterion test consisting of a series of multiple-choice items pertaining to the known life experiences, attitudes, etc. of a subject called David. The sets of data utilized were selected in part to study the contribution of certain types of information to clinical diagnosis and appraisal.

METHOD

Criterion Test

The task posed for the several groups of judges was that of providing correct answers to three types of questions, some dealing with David's current interests and attitudes, others dealing with certain of his "characteristic" behavior patterns, and still others dealing with his responses to specific recent life situations. These questions were presented in the form of four-alternative multiple-choice items, each set of alternatives being preceded by a statement providing pertinent facts for, and setting the context of, the question. Following is an example:

On one occasion two foreign students who had completed their studies and were leaving the school, arranged a small party for their friends. Some mild spirits were available to those interested, although as nearly as can be ascertained nothing stronger than wine was served. In this school, opinion is divided on the matter of social drinking. Some students object strongly, others are more tolerant of moderate social drinking. Members of both groups were present at the party.

a. At first, David was a little uncomfortable, but

¹ For invaluable assistance in carrying out various phases of this study the writer is deeply indebted to Sarah Counts, D. W. Fiske, Irving Leiden, Joanne Powers, Rae Shifrin, Saul Siegel, and Alvin Winder.

² This study was carried out while the author was at the University of Chicago.

he joined in, swapped stories with the rest, and drank.

b. After a couple of drinks he began a rather unconvincing act of being slightly intoxicated, began to slur his speech, walk uncertainly, etc.

c. He conspicuously refused a proffered drink, and shortly after reprimanded the host for serving liquor to such a gathering.

d. He remained throughout the party, but made careful note of the persons who accepted drinks and those who told questionable jokes and stories, and subsequently dropped them from his list of social acquaintances.

The material for these items came principally from three sources: interviews with David, interviews with his wife and other associates, and direct observation by members of the research staff. David participated in a total of approximately nine hours of personal interviews with the investigator. During these sessions, which were conducted as relatively unstructured history-taking interviews, he was encouraged to talk freely about his marriage, his childhood, his present work and aspirations for the future, etc. Aside from occasional questions setting a new topic or seeking further elaborations, there was minimal intrusion by the interviewer. These sessions and all other interviews in this study were recorded on wire for subsequent reference.

A second, though no less important, source of information was David's associates—his wife and a number of close acquaintances, both male and female. Upon request, David provided the names of a number of acquaintances with whom he was in frequent contact, persons whom he had no objections to our interviewing. From this list, seven or eight persons were interviewed, some only once or twice, several others three or four times over a two-month period. Each of the interviewees was given a careful explanation of the nature of the project and was apprised of the fact that David had given his permission for such interviews to take place. (David himself did not know which persons were to be called for interviews.) These interviews focused largely on current events in David's life.

David's wife was interviewed twice, for a total of approximately three hours. Her interviews were moderately structured, since the main objective here was to gather indirectly information that might corroborate or qualify statements made by her husband.

Direct observation yielded relatively little useful information because the periods of observation were few and brief and because there was no way of obtaining corroboration of the single observer's reports. Yet, it was possible by this method to obtain clues to other appropriate item material.

From this collection of observational notes and

recordings, an initial collection of 54 items was constructed. The present analysis, however, deals with only 37. Eight items were dropped because they differed slightly in two versions of the test, and nine others were omitted because all persons in each of the four groups taking the test failed to choose the correct alternative. Consequently, they were thought either too difficult or patently misleading in structure. In assembling the test, some attempt was made to include items pertaining to a broad range of behaviors; however, there was no conscious or planned attempt to include equal numbers of items indicative of good adjustment and poor adjustment—an outcome which, though subsequently important, must be regarded as somewhat fortuitous.

Critical Data

The designation "critical data" refers to the different types of specialized information provided to the several judging groups. Four different classes of data were considered: a collection of biographical facts, most of them provided in the context of the criterion test items; observations of the subject in a series of role-playing situations, the subject's Rorschach protocol, and a battery of psychological tests.

Role-playing situations. Two weeks prior to the day on which David participated in the role-playing sessions, a group of graduate students participating in the study—five of whom were later to serve as judges in the role-playing session—examined the criterion test and designed nine situations which they considered as likely to elicit information pertinent to the questions raised in the criterion test. None of the situations was "tailored" to particular items; rather, each was designed to provide information from which a variety of inferences might be drawn. Two of the situations were small-group discussions (e.g., David as a new, young superintendent of a boys' correctional school discussed school policy with veteran staff members). The remaining seven situations involved only the subject and a standard role-player. In four of these seven situations the standard role-player was a male, in three, a female. In the four all-male situations, David's relationship to the other person was varied so that once he was a subordinate (a news reporter being pressed by the editor to meet a deadline), once in an ambiguous position (the young and inexperienced institutional superintendent talking to a long-time staff member with strong political connections), and twice a peer (a potential client discussing with an architect plans for buying property and building a house, and discussing philosophies of life with a carefree artist whom he met at a cocktail party). In the male-female situations he was once a peer (to the wife of a friend who seeks advice on a marital problem), once a superior (to a young teacher whose behavior both in school and out has caused "some talk" in the small community where he is superintendent of schools), and once in a quasi-dependent position (that of a convalescing patient with an oversolicitous, mildly dominating nurse).

In all, David interacted with five different people, three women and two men. Two of the women were five to seven years older than David, who was about 26 years of age, and one was about three years younger. One of the men was about six years older than David,

the other about his own age. These standard role-players were picked for their ability to adapt easily to these particular roles. All of them were instructed as to the general character of the situation they were to try to create, but were left free to implement these directions in their own fashion as the situation developed. The role-players were given two opportunities to run through their roles prior to the day on which the sessions were held.

Psychological tests. Prior to the beginning of the history-taking interviews, a day or two after David had been given an account of the nature of the study and had agreed to participate, he was given a battery of psychological tests, including the Wechsler-Bellevue, Rorschach, TAT, a sentence completion test, a picture-drawing test (man, woman, building), the Allport-Vernon Scale of Values, the Guilford-Martin inventories (GAMIN and STDCR), and a 60-item word-association task. The Rorschach and TAT responses as well as the inquiry on the former were wire-recorded and transcribed. Judges using the Rorschach received hectographed copies of the original responses and inquiry, together with an accurately marked location chart and a brief description of the subject's behavior while taking the test. Judges using the entire battery received these same materials plus duplicate copies of the other test records and photostatic copies of the picture-drawings.

Biographical data. All subjects were told David's age, race, religion, marital and educational status, and present occupation. In addition, wherever relevant, biographical material was contained in the stems of the criterion test items. Hence, all judges knew certain salient facts about his military history, his career objectives, his extracurricular interests and accomplishments in college, his financial problems, etc.

Judging Groups

Role-play judges. The persons who observed David in role-playing situations (RP judges) included five advanced graduate students in clinical psychology who had been participating in a seminar on the analysis of role-playing behavior, and four psychiatrists, none of whom had any previous experience with role-playing situations in this type of setting.

Rorschach judges. Seven persons (Ror. judges), five of whom were generally regarded as experienced Rorschach diagnosticians and who function professionally in that capacity, evaluated David's Rorschach. Several of them were "authorities" in the sense that in addition to years of clinical experience with the instrument, they had published on and taught advanced courses on Rorschach testing. The two less experienced persons were advanced psychological interns, functioning in the setting of a mental institution.

Test battery judges. The group concerned with the test battery (TB judges), consisted of 39 graduate students at various stages of training in clinical psychology. All were second- or third-year interns in the Veterans Administration clinical psychology training program.

Control groups. Two control groups were utilized. The first ("Guess" judges) was a group of 12 psychologists enrolled in an advanced Rorschach workshop.

These judges studied the criterion test and made their item selections solely on the basis of the related facts reported in the test items.

A second group consisted of 21 student nurses. For this group a special version of the test was prepared in which all items were identical to the standard criterion test except that a *different man's name was used in each item*. The nurses were told that the items constituted a test of their knowledge of male behavior patterns and were based on the actual behaviors of a group of white, married (no children) theology students between the ages of 25-27, the group from which David was drawn. By this means it was possible to neutralize the cumulative biographical information about David otherwise available in the criterion test. The nurses were asked to study each item and to try to guess which behavior might be most probable for a man with the characteristics indicated.

Procedure

At intervals varying from two weeks to three days before receiving the critical data, Role-play and Rorschach judges were given a copy of the criterion test and were asked to answer the items by using whatever cues they could obtain from the biographical data available in them. The objectives of this assignment were twofold: (a) to acquaint judges with the exact types of questions they would be required to answer subsequently on the basis of the critical data, and (b) to obtain information about the prevailing stereotype held by judges for such a person as David. Scores earned at this stage, before exposure to the critical data, hereafter are referred to as stereotype scores. All Ror. judges and five of the RP judges completed the stereotype assignment only two or three days before the critical information was provided. The five RP judges who participated in designing the role-playing situations completed the stereotype tasks about two weeks before the day on which they saw David perform in these situations.

The TB group did not do the stereotype task; only one day of this group's time was available for the study, and the entire day was required for their studying the test battery and making the final set of responses. The Guess group, on the other hand, was asked to do only the stereotype task.

In point of time, David participated in the role-playing situations about four months after first consenting to participate in the project. On the assigned day, he was brought to the laboratory and was given an explanation as to the nature of the role-playing situations as well as a brief and very general description of the task confronting the judges. It was explained to him that all the judges and all the standard role-players except one—a person he had met briefly through the investigator—were complete strangers to him, persons it was unlikely that he would encounter outside of this research situation.

RESULTS

Differences in Over-all Accuracy

The mean and range of scores for each of the groups of judges is reported in Table 1. Sig-

nificance of differences in performance between the groups was tested by comparing means of the sum of ranks (1) for the several groups participating in the stereotype stage and in the final stage. In the former instance $H = 4.099$, with $df = 2$, $.10 > p > .25$. In the latter instance $H = 2.739$, with $df = 2$, $.25 > p > .50$. Significance of differences between stereotype and final scores for RP judges and for Ror. judges was tested by Wilcoxon's T method for ranks (1). In neither case were the differences significant. It is possible to conclude, therefore, that none of the three types of critical data improved accuracy scores (with respect to the specific type of criterion employed in this study) beyond the level achieved by study of biographical facts alone.

Because of the similarity in mean scores for all six judging tasks, general characteristics of the criterion test were examined to determine whether difficulty of the test imposed a uniformly low ceiling on judge performance. It was found that 14 of the 37 items were correctly identified by 75% or more of the judges in at least one of the six subgroups of judges; 21 of the items were correctly identified by 67% or more of the judges in at least one of the groups, and 29 items were identified correctly by 50% or more of the judges in at least one group. In the remaining eight items, the correct alternative was chosen by at least 40% of the judges in one or more groups. A total of 25 items met the double criterion of having been correctly identified by at least 50% of the judges in at least two of the six subgroups, or by at least 67% of the judges in any one judging group. It was evident from this inspection that the similarity of means was not a resultant of the criterion test's containing a pool of easy items which most judges answered correctly under all

TABLE 1
MEAN AND RANGE OF SCORES ON CRITERION
TEST BY FIVE DIFFERENT JUDGING GROUPS

Judges	N	Stereotype		Final	
		Mean	Range	Mean	Range
Role-play	9	14.56	10-19	14.78	9-23
Rorschach	7	15.00	11-20	13.71	11-16
Test battery	39	—	—	15.49	9-22
Guess	12	12.00	6-19	—	—
Nurses	21	10.81	8-14	—	—

circumstances and another pool of difficult items which most judges failed.

Differential Accuracy

Performance of the judging groups was analyzed in two different ways in order to ascertain whether there were significant differences in accuracy on different parts of the test. These analyses had to do on the one hand with the types of tasks faced by the judges, and on the other hand with the degree of maladjustment implied in the item alternatives.

Item types. Regarding the task approach, the items making up the criterion test posed three presumably different types of questions for the judges. One group of 17 items required judges to postdict David's behavior in specifically described situations; a second group of 9 items called for the specification of general characteristics of his behavior, and the third, containing 11 items, pertained to certain of David's interests and attitudes. Mean scores for RP and Ror. groups on these three types of items are reported in Table 2.

Differences in performance on the three types of items at a given stage were tested by means of analysis of variance of ranks in a twofold classification, based on the proportion of successes in each category by each individual. For RP judges at the stereotype stage, chi square was 10.0, with $df = 2$, $p < .01$. For Ror. judges at the stereotype stage, chi square was 6.7, with $df = 2$, $.02 < p < .05$. No significant differences were found at the final stage.

Differences in accuracy within item types at two different stages, i.e., between stereotype and final stage, were tested by Wilcoxon's T

method. RP judges showed no significant change in accuracy on any of the three types of items; Ror. judges, however, showed a significant decrease ($p = .05$) in accuracy on postdiction items from stereotype to final stage.

Maladjustment index. The second analysis was aimed at investigating differential accuracy in relation to a particular set of hypotheses or attitudes which judges might form about David as a consequence of exposure to particular kinds of data. Previous studies (2, 3) suggested that individuals relying primarily or exclusively on projective tests for their critical data tended to overestimate a subject's maladjustive propensities. It was hypothesized, therefore, that the Ror. judges would tend to overestimate the subject's maladjustive tendencies and hence choose those response alternatives most characteristic of a maladjusted person.

To test this possibility, the 148 alternatives in the criterion test were first rated on a seven-point scale of adjustment-maladjustment by five judges working independently. Interjudge agreement ranged from a correlation of .53 to .71, with a median of .69. The median of the ratings for the correct alternative in each item was used to classify the items according to the degree of adjustment or maladjustment implied in the correct alternative. Three groups of items were delineated, one consisting of the 12 items whose correct alternative received the highest median ratings on adjustment, another of the 12 items whose correct alternative received the highest median ratings on maladjustment, and a third group containing

TABLE 2
MEAN ACCURACY SCORES OF ROLE-PLAY AND RORSCHACH JUDGES ON THREE CLASSES OF ITEMS AT STEREOTYPE AND FINAL STAGES

Item Type	N	Stage	Mean Accuracy Scores	
			Role-play	Ror-schach
Postdiction	17	Stereotype	8.00	7.43
		Final	7.11	5.43
Characteristics	9	Stereotype	2.00	2.56
		Final	3.00	3.57
Interests and attitudes	11	Stereotype	4.56	5.11
		Final	4.57	4.77

TABLE 3
MEAN ACCURACY SCORES OF NURSES, ROLE-PLAY, RORSCHACH, AND TEST BATTERY JUDGES ON ADJUSTIVE, INTERMEDIATE, AND MALADJUSTIVE ITEMS

Judges	Stage	Item Type			Total Test
		Adj.	Int.	Mal.	
Nurses	Stereotype	5.82	3.36	1.68	10.81
	Role-play	6.78	5.00	2.78	14.56
	Final	7.89	5.33	1.56	14.78
Rorschach	Stereotype	5.57	5.00	4.43	15.00
	Final	3.57	3.86	6.29	13.71
	Test battery	5.13	5.74	4.62	15.49

the 13 intermediate items. Responses of the RP, Ror., Nurses, and TB judges were tallied to yield subscores for these three classes of items. Group means for these accuracy subscores are reported in Table 3.

Differences between scores at stereotype and final stages for a given group of judges were tested by Wilcoxon's *T* method. Differences between groups of judges for a given type of item at either stereotype or final stage were tested by the method of analysis of variance of ranked scores. Significance levels of the observed differences are reported in Table 4. For RP judges, difference in performance on adjustive items as compared with maladjustive items was significant at both the stereotype stage and the final stage; for Ror. judges, such a difference was found to be significant only at the final stage. The three groups of judges participating in the stereotype task differed significantly in performance on the maladjustive items, but not on adjustive or intermediate items. At the final stage, however, after exposure to the critical data, the three groups of judges who completed that task showed significant differences in accuracy on both adjustive and maladjustive items.

Finally, RP judges showed a significant decrease in accuracy on maladjustive items as they advanced from the stereotype to the final task, whereas Ror. judges showed a significant decrease in accuracy on adjustive items from the former to the latter stage.

The major features and interrelationships of these differences as they pertain to RP and Ror. judges are quickly grasped from Fig. 1. Since each of the four alternatives to an item was assigned a maladjustment index by the

rating procedure described earlier, it was possible to determine the mean maladjustment index of the preferred choices of all the members of a judging group for each item and, hence, for each of the three major groups of items. On a given item, for example, the mean maladjustment index was arrived at by multiplying the frequency of selection of a given alternative by the maladjustment weight earlier assigned to that alternative and then determining the mean of these values for the four alternatives in the item. By extension, a similar mean maladjustment index could be determined for any group of items. Thus, the mean maladjustment index for all the choices by RP judges at the stereotype stage for the 12 adjustive items was 3.3, for the intermediate items 3.9, and for the maladjustive items 3.4. By comparing the profiles for the two groups at the stereotype stage with those for the final stage, it becomes apparent that the net effect of the critical data provided to the two groups after the stereotype stage was to increase differences between them. RP judges, who tended to choose the more favorable alternatives at the stereotype stage, simply increased that tendency, whereas Ror. judges, who tended to choose the somewhat less favorable alternatives at the stereotype stage, increased that tendency at the final stage.

Figure 2 sheds further light on judges' preferences. The alternatives for each item were ranked according to the magnitude of the mean of the maladjustment weight assigned to each of them by the five independent judges. Then choices of Ror., RP, and TB judges at the final stage were examined to determine frequency of preference for items of different

TABLE 4
P VALUES OF SIGNIFICANT DIFFERENCES BETWEEN RANKS OF ACCURACY SCORES ON SUBSECTIONS OF THE TEST AT STEREOTYPE AND FINAL STAGES

	Stereotype		Final		Stereotype vs. Final			
	Comparison	<i>p</i>	Comparison	<i>p</i>	Role-play	<i>P</i>	Ror-schach	<i>p</i>
Judges								
RP								
Ror.	Adj. > Mal.	.02	Adj. > Mal.	.01				
Items	Adj. > Mal.	—*	Mal. > Adj.	.05				
Adj.								
Int.	RP > Ror. > G	—	RP > TB > Ror.	.001	S < F	—	S > F	.05
Mal.	RP > Ror. > G	—	TB > RP > Ror.	—	S < F	—	S > F	—
	Ror. > RP > G	.001	Ror. > TB > RP	.001	S > F	.01	S < F	—

* Not significant.

of background situational information for each item. The attempt to assess the effect of this type of information on judges by having them make preliminary responses may foster certain "sets" which influence their appraisal of data at later stages.

Aside from general familiarity with the "halo-effect," not a great deal is known about the influence of the initial set or attitude of diagnosticians on the interpretation of data. In an earlier study (2), two teams of judges, simultaneously observing the same two groups of subjects in role-playing situations, exhibited systematic differences in their evaluations, depending upon their relationship to the subjects. In the present instance, another factor apparently influenced initial set. The significant group differences between RP and Ror. judges in evaluation of the biographical information may be accounted for by differences in attitude toward the data at the stereotype stage. It is possible that mere knowledge that their next task would demand further blind appraisal based on the Rorschach alone heightened the sensitivity or the responsiveness of Rorschach judges to adverse signs in the biographical data. The similarity of these two sets of findings suggests the need for more thorough studies of the relationship between set and judgment in diagnostic situations.

A much more basic difficulty with the present type of criterion arises in the selection of materials for item construction. The occurrence of a balanced collection of adjustive and maladjustive items in the present criterion test was not intentional; indeed, the maladjustment ratings were gathered after the test had been administered to RP and Ror. judges. One of the implications of this array of items was that David usually behaved in a manner characteristic of moderate adjustment, whereas fuller acquaintance with and a careful accounting of all of David's behavior over a long period of time might indicate that the true distribution is more skewed. Hence, interpretation of the scores earned by the several groups must be made with caution, for if the *true* distribution for David is in fact skewed in one direction or another, then the present results on differential accuracy between groups are misleading. It should be pointed out, however, that the very considerable amount of data collecting

that lay behind the construction of the criterion test was undertaken precisely in order to sample a very large collection of life situations, and that viewed against this background, David appeared to have neither more nor less than his share of personality assets and liabilities.

Of far greater interest than the scores themselves are the circumstances that produced essentially similar scores in groups of judges who appear to have held quite different conceptualizations of the subject.

Kelly and Fiske (4), Soskin (3), and others have reported results indicating that access to additional data beyond some optimal amount does not increase the accuracy of appraisal or the prediction of human behavior. The present study clearly indicates that such a conclusion reflects only part of the story. Here, two groups of judges presented with identical data—biographical facts—formed slightly different conceptualizations of the subject. When each of the groups was then provided with a different type of additional data, one group exhibited only a slight change in its initial conceptualization, whereas the other exhibited a marked change.

The divergence occurring in this second stage must be presumed to have occurred because the low-change (Role-play judges) group found the additional data to be essentially consistent with its initial hypothesis, i.e., the additional data revealed little which was not already known or inferable from the biographical data; whereas the high-change group (Rorschach judges) found its additional data to be in large measure inconsistent with the initial hypothesis, i.e., the judges had obtained a considerable amount of new information to which they attached a high degree of credibility.

Keeping in mind that the two groups differed significantly on maladjustive items at the initial stage, it appears that for each group initial conceptualizations were effected through an integration of certain selected salient consistencies in the data, and that contradictory data were regarded by the judges as instances of tolerable inconsistency and, hence, were ignored.

Apparently, once formed, such a relatively undifferentiated conceptualization is main-

tained regardless of the amount of additional data subsequently provided, so long as the additional data yield the same relative proportions of confirmatory and contraindicative information. However, when a sufficient mass of new information which the judge regards as highly credible, relevant, and contraindicative becomes available, the initial formulation is likely to change. Much of the material previously ignored or discounted may then be integrated into the new conceptualization, while evidence previously regarded as supporting a different conceptualization may be discounted.

This appears to have been the experience of Rorschach judges in the second stage. A sufficient amount of relevant, credible, contraindicative information appeared in the Rorschach data to warrant abandoning the initial conceptualization in favor of an essentially new one. It need scarcely be pointed out that both the relevance and credibility of certain of the new information, at least with respect to the present criterion, was somewhat overestimated by the majority of the Rorschach judges, inasmuch as they abandoned features of the initial conceptualization which were essentially sound even though based on biographical facts alone.

In view of the consistency of preference trends among members of a given group of judges, it appears that, wittingly or unwittingly, the fundamental ingredient in the conceptualizations of the majority of judges was a gross judgment as to David's adjustive vs. maladjustive propensities, and this judgment was brought to bear in the majority of decision situations in a fairly consistent manner.

It seems reasonable to assume that whenever a very large number of decisions is called for, each of which requires the evaluation of a substantial number of subtly related facts, the most parsimonious solution is to predict from a general characteristic rather than to attempt to integrate a large number of contingent probabilities for each separate decision.

In general, the Rorschach findings are consistent with results obtained in two previous studies (2, 3) in which personality appraisals based solely or primarily on Rorschach information were found to err quite consistently in the direction of overestimating the degree of maladjustment.

The performance of Role-play judges is more perplexing. An implicit basic hypothesis in this study was that situational behavior is largely a function of ego processes whose characteristics are only poorly represented in the perceptual response to the Rorschach cards. To assess these processes it is necessary to create test conditions which elicit them, and it was felt that role-playing situations would accomplish precisely this end. The present results, however, suggest that experienced clinical observers, given a relatively long period of observation of a variety of interpersonal situations, were unable to make the discriminations that had been expected of them. The striking similarity in pattern of preferences exhibited by the nurses (who had only the grossest general information) and the Role-play judges commands attention. It suggests that for all the specialized information available to them, the Role-play judges as a group, both before and after obtaining the specialized information, entertained essentially the same expectations of the subject's behavior as would a group of unsophisticated young women operating solely on the basis of their general social experience. The results suggest the following possibilities:

1. Role-playing behavior leaves the defenses sufficiently intact to preclude the possibility that most judges could detect maladjustive tendencies; or,
2. The task assigned to judges in this study was so complex as to oblige them to make most decisions on the basis of a single over-all impression; or
3. The majority of judges developed a positive affective response toward the subject which caused them to evaluate his behavior with a selective bias.

SUMMARY

In a diagnostic task where groups of judges were called upon to specify certain behaviors, characteristics, and interests and attitudes of a subject (a) on the basis of biographical facts alone and then (b) on the basis of either observation of role-playing situations, interpretation of a Rorschach protocol, or study of a battery of objective and projective tests:

1. The five groups of judges involved did not differ significantly in over-all accuracy regardless of the type of information on which their judgments were based.
2. Judges who observed the subject in role-playing situations and judges who studied the subject's Rorschach protocol showed no greater

accuracy after exposure to these types of data than had been achieved by study of biographical facts alone.

3. Information obtained from observing the subject in nine role-playing situations appeared merely to confirm the conceptualization which a group of judges had formed from study of biographical data.

4. Information obtained from the Rorschach protocol increased the tendency of a group of judges to choose items characteristic of maladjustive behavior beyond that which was evident when the group studied only biographical data.

5. As compared with observers of role-playing, Rorschach interpreters showed a significantly greater tendency to change their selections after exposure to the critical data;

in neither group, however, was amount of change significantly related to increase or decrease in accuracy.

6. Greatest flexibility of characterization of the subject was exhibited by a group of judges who based their selections on a diversified battery of objective and projective tests.

REFERENCES

1. WALKER, HELEN M., & LEV, J. *Statistical inference*. New York: Henry Holt, 1953.
2. SOSKIN, W. F. Frames of reference in personality assessment. *J. clin. Psychol.*, 1954, **10**, 107-114.
3. SOSKIN, W. F. Bias in postdiction from projective tests. *J. abnorm. soc. Psychol.*, 1954, **49**, 69-74.
4. KELLY, E. L., & FISKE, D. W. *The prediction of performance in clinical psychology*. Ann Arbor: Univer. Michigan Press, 1951.

Received September 19, 1957.

THE SOCIAL PSYCHOLOGY OF RORSCHACH VALIDITY RESEARCH

LEON H. LEVY AND THOMAS B. ORR

Indiana University

RECENT years have seen an increasing appreciation of the possible contribution of the personality of the examiner to the variance of projective test protocols (1, 4, 5, 6, 7, 10, 12). While many of the studies concerned with this problem may be subject to methodological criticism (8), they are all fairly consistent in their finding of a significant relationship between various personality traits of the examiner and the scores and interpretations of tests administered by him. Until the nature of the mechanisms mediating these relationships is fully understood, generalizations about the utility of projective tests *per se* seem extremely hazardous. For example, when Investigator A fails to replicate the findings of Investigator B, are his results a function of his personality or of the instrument or theory in question? We have no way of apportioning the variance.

Such problems may be construed as falling within the broader purview of the sociology of knowledge (9), where the ideas and information of individuals or groups are related to their particular social status and values. Within such a context, one may ask to what extent the outcome and interpretation of research is conditioned by the circumstances under which it is instigated and conducted. This question does not refer to methodological issues but, rather, to the needs, interests, and commitments of the individual researcher and the institutions with which he is affiliated. Raising this query implies no derogation of either the scientific integrity or rigor of the researcher; it merely asks whether what may be called the social psychology of research is an important consideration in evaluating any substantial body of investigation or a potentially fruitful area for further investigation.

Psychologists, like other scientists, rely upon the logic of statistical inference and experimental control in attempting to divorce their findings from their own personal beliefs and prejudices. If they are justified in this reliance, then the outcome and interpretation of research concerning any controversial issue

should not be related to the theoretical or institutional commitments of the researcher. Should this not be the case, important methodological and substantive considerations would necessarily follow. As a preliminary foray into this area, we have chosen to analyze in a fairly gross way the social psychology of Rorschach validity research.

The Rorschach validity literature is particularly well suited to these purposes. First, it is extensive. Second, it is far from univocal in its findings. Third, it is performed by researchers with varying institutional affiliations. And last, it seems possible to make certain plausible inferences regarding the dominant needs and interests associated with these institutional settings and the kinds of research findings that might be most consistent with them. The problem then is to determine whether such a relationship does in fact exist.

We shall be concerned with the relationships between three dichotomized variables in this study: (a) the type of institution in which the research was conducted, whether academic or nonacademic; (b) the type of validity study, whether construct or criterion-oriented (3); and (c) the outcome or interpretation of the research, whether it was favorable or unfavorable with regard to the validity of the Rorschach.

From a consideration of the presumably dominant interests and concerns of individuals affiliated with academic and nonacademic settings several hypotheses were generated. To begin with, it seems reasonable to assume that the investigator who owes his primary allegiance to an academic institution is more likely to be interested in the theoretical aspects of a diagnostic instrument, i.e., its construct validity, than with its practical aspects or its criterion validity. This is not to say of course that the academician is unconcerned with the practical utility of tests—only that he is more likely to consider validity as contingent upon a thorough theoretical understanding of the instrument. The researcher working under the auspices of a nonacademic institution, on the

other hand, is in a setting where a diagnostic device must pay its way in the making of day-to-day clinical decisions. Hence, he is more likely to be concerned with its criterion validity than with its construct validity. Again, it is incorrect to suggest that such individuals have no theoretical interests or perform no theoretical research. But research in a service setting is more frequently considered a luxury item, and, if conducted, it is more likely to be of a practical than theoretical nature. The first hypothesis, then, may be formulated as follows: Rorschach validity research performed in an academic setting is more frequently construct-oriented than criterion-oriented, whereas the reverse holds true for research conducted in a nonacademic setting.

In line with the dominant kinds of enthusiasms postulated for each type of setting, it seems reasonable to assume that research results consistent with these enthusiasms would be more highly valued than those not consistent with these interests. Thus, the theoretically oriented psychologist may be more concerned with obtaining positive results in a construct validity study than in a criterion validity study, whereas the reverse holds true for the practically oriented psychologist. Accordingly, the second hypothesis asserts that in studies emanating from academic institutions there is a higher proportion of favorable results obtained in construct validity studies than in criterion validity studies, whereas the reverse holds true for studies performed in nonacademic settings.

As a corollary of the second hypothesis, a third may be formulated: A larger proportion of construct validity studies performed in academic settings than in nonacademic settings is favorable with respect to Rorschach test validity, whereas the reverse relationship holds for criterion validity studies.

Finally, implied in these three hypotheses is a fourth hypothesis that the three variables—institutional setting, type of validity study, and interpretation—are interrelated.

These four hypotheses are all predicated upon the assumption that a relationship exists between the institutional affiliation of a researcher or the institutional aegis under which research is conducted, the kind of research performed, and the probability of obtaining

certain types of outcomes or interpretations of data. To the extent that the null statements of these hypotheses may be rejected at statistically reliable levels, this assumption gains in credibility.

METHOD

On the assumption that the bulk of Rorschach validity studies is published in the *Journal of Abnormal and Social Psychology*, *Journal of Clinical Psychology*, *Journal of Consulting Psychology*, and the *Journal of Projective Techniques*, each article which appeared in these journals during the years 1951 through 1955, inclusive, was inspected to see if it met the criteria for inclusion in this study. The year 1951 was chosen as a starting point because Cronbach's (2) statistical critique of Rorschach research methodology appeared two years earlier. Given a maximum publication lag of two years, research from 1951 on might be expected to show greater methodological sophistication than that done previously. No attempt was made, however, to check the validity of this expectation.

Criteria for Selection

To be selected, a study had to test at least one hypothesis, explicit or implicit, relevant to the interpretation or to some application of the Rorschach test. While studies of reliability, internal consistency, administration and scoring techniques, examiner influences, certain normative studies, etc., may have some implications for the validity of the test, they were not considered here as evidence on validity. Including some of these studies would have made the categories of validity closer to Cronbach and Meehl's (3) definitions but would have introduced additional subjectivity in selection and classification. Studies which investigated widely used principles of Rorschach interpretation, such as "color shock," were included even if the Rorschach plates had not been used, so long as the purpose had been to provide information specifically relevant to Rorschach usage. It was not required that an article consider the Rorschach only; some ambitious studies could conceivably have been divided into a number of separate articles, each considering hypotheses relevant to a single instrument. Only the Rorschach hypotheses were considered, however, in classifying such a study.

At least one statistical test that would allow a statement as to the probability of the obtained results was also required. This criterion excluded purely expository articles, case studies, and a few other articles that offered scores or frequencies without any indication as to their significance. A further requirement, naturally, was that the studies be reported in sufficient detail to be subsequently categorized on the variables considered here.

One hundred sixty-eight studies were found that met the indicated criteria. Each was coded for each of the three variables studied, together with additional identifying information on individual cards. All cards were placed in a common pile as they were filled out and were not referred to until all articles had been classified.

Definition of Categories

The institutional variable. Institutional affiliation was dichotomized along the academic-nonacademic axis. Studies were classified on the basis of the affiliation of their senior author. If the location listed under his name was a college or university, including schools of medicine, the setting was considered academic. All other affiliations were considered nonacademic. Whenever a footnote indicated that the study had served as the basis for an academic degree, the setting was classified as academic even though the author was associated with a nonacademic institution at the time of publication. Similarly, if a footnote indicated that the entire study had been conducted under a previous affiliation, the classification was made on the basis of the earlier setting. When the senior author listed more than one affiliation, the one which agreed with the affiliations of the majority of his co-authors was used. In the case of a single author with more than one affiliation, the first listed was used as the basis for classification.

Type of validity study. Studies were classified as bearing upon either the construct validity or the criterion validity of the test. Cronbach and Meehl's (3) definitions were used, with both predictive and concurrent validity studies subsumed under the heading of criterion validity. Any study in which the variable related to the Rorschach was considered to be of immediate value in practical decision-making was placed under the criterion validity classification. A construct validity study was considered one in which the non-Rorschach variable was of primarily theoretical interest and could be used only indirectly, if at all, in the clinical decision-making process. This distinction is essentially the same as that Cronbach and Meehl make between studies where the primary interest is in predicting the criterion and the test behavior is only of minor or secondary interest and those where the major concern is with the test itself and the criterion is chosen on the basis of theoretical considerations. In those studies which seemed to have both criterion and construct validity implications, classification was based on what appeared to be the major concern of the author.

The outcome or interpretation. On research outcome, studies were classified as favorable or unfavorable with regard to Rorschach validity on the basis of the investigator's conclusions from his own data. In those instances where several hypotheses or several tests of the same hypothesis were considered and where not all of them reached levels of statistical significance, the investigator's evaluation of his findings still served as the basis for classification. Where the author made no clear statement regarding the interpretation to be placed on his data, a judgment was made by the present authors.

Clearly, investigators differ in the stringency of the criteria which they apply in the evaluation of their findings. However, for the purposes of the present study it seemed that the appropriate datum was the author's own evaluation. Thus, the question remains of whether, granting such individual differences, they are less than or greater than any differences among types of institutional affiliation.

RESULTS

Interjudge agreement on the classification of 40 randomly selected studies judged independently by a second person on each of the three variables was 98% on the institutional variable, 93% on the type of study variable, and 90% on the outcome variable. While these are all sufficiently high to warrant confidence in the objectivity of the classifications which form the basis of this study, it is interesting to note that none was completely unambiguous.

All hypotheses were tested by means of chi square analyses applied to various combinations and rearrangements of the data presented in Table 1. Table 2 presents a summary of the results of these analyses. In all instances, the tests were based on one degree of freedom. The over-all chi square testing the fourth hypothesis and applied to the entire table was performed by Norton's (11) method for calculating chi square for complex contingency tables.

From Table 2 it will be seen that the first hypothesis received support at a high level of confidence. Apparently, the kind of validity study done is related to the setting in which it is performed. The second hypothesis is supported only in the case of studies performed in an academic setting; hence it must be considered only partially confirmed. In the case of research performed in academic institutions, there is a significant relationship between the kind of validity study undertaken and the probability of favorable or unfavorable results, but there is no evidence for such a relationship in studies issuing from nonacademic institutions. In an academic setting, the chances are better than two to one that a construct validity study will yield favorable results vis-à-vis the Rorschach, while the odds are almost two to one that a criterion validity study will yield negative results.

The test of the third hypothesis reached significance at the .05 level for both the con-

TABLE 1

DISTRIBUTION OF STUDIES AMONG VARIABLES OF SETTING, TYPE OF STUDY, AND OUTCOME

	Academic		Nonacademic	
	Construct	Criterion	Construct	Criterion
Favorable	51	12	14	19
Unfavorable	22	23	14	13

TABLE 2

SUMMARY OF RESULTS OF CHI SQUARE TESTS FOR FOUR HYPOTHESES

Hypothesis	Comparison	χ^2	p^*
1	Type Study \times Setting	7.04	.01
2	Results \times Type Study (Academic)	12.32	.001
	Results \times Type Study (Nonacad.)	.53	N.S.
3	Results \times Setting (Construct)	3.48	.05
	Results \times Setting (Criterion)	4.23	.05
4	Results \times Type Study \times Setting	7.71	.01

* All probabilities except that for Hypothesis 4 are for a one-tailed test because the direction of differences was predicted in each case.

struct and criterion validity studies, thus supporting the contention that for a given type of validity study, the probability of a positive or negative outcome varies with the kind of setting in which it is performed. Similarly, the fourth hypothesis, asserting the interrelationship of the three variables, received support at the .01 level of significance.

DISCUSSION

The results clearly support the hypotheses concerning the interrelationship among the institutional setting of research, the kind of research done, and the type of outcome. In turn, they strongly suggest that in approaching any body of research dealing with a particular issue, one must inquire beyond the "box score" in order to evaluate the findings. There *does* appear to exist a social psychology of research.

The relationship between institutional setting and type of study has serious implications. If it were found that construct validity studies generally had a better chance than criterion validity studies of producing positive results,

there would be no cause for alarm except among those using the Rorschach to decide between such alternatives as brain damage or no brain damage, psychotic or neurotic, to shock or not to shock, to treat or not to treat, etc. Such a finding would imply that the test is not ready for such tasks or that we have not yet learned how to construe these alternatives in terms of the constructs measured by the test, assuming their relevance. But when the probability of a favorable outcome in a construct validity study drops from 70% in academic to 50% in nonacademic research, and the probability of a favorable outcome in a criterion-oriented study rises from 34% to 59% from one setting to the other, it gives one pause in the marshalling of either pro- or anti-Rorschach evidence!

The categories used in the present study were admittedly gross, and one may ask what one would find if the research were more finely classified in terms of problems investigated, techniques and procedures used, etc. Of no small interest and importance is the study of the characteristic research output of particular individuals. Unsystematic observation certainly suggests that there are persons who consistently report either positive or negative findings with regard to the Rorschach.

One may also be concerned with variations in quality of research, a problem that certainly must be attacked eventually. The articles reviewed all purported to follow accepted experimental and statistical procedures, but the findings clearly indicate the presence of some sort of bias. Either some of the studies were methodologically weak, or present statistical concepts need further scrutiny. For example, does the .05 level of significance mean the same thing when used as the criterion for rejection of the null hypothesis in a construct validity as opposed to criterion validity study executed in an academic setting? On the basis of the present data, it seems clear that this is not the case. Other things being equal, a construct validity study has a 70% chance of producing significant results at this level or better, whereas a criterion validity study has only a 34% chance of doing so. Surely some allowance should be made for such discrepancies in deciding whether or not a null hypothesis may be rejected and in stating the confidence with which this decision is made.

But, as Mannheim (9) points out, to accept the present findings as a basis for discounting certain studies because of their institutional origin would be in serious error. The reported results not only pose a barrier in the way of those who would uncritically gather their support or intellectual ammunition; they may also hold the germ of greater insight into human behavior. Attempting to manipulate probabilities so as to compensate for base rates is only one solution to the problem raised here. A more challenging and potentially more rewarding approach lies in the intensive study of the behavior of the researcher himself. For, intentionally or not, he seems to exercise greater control over human behavior than is generally thought. By the investigation of the process of research as conducted in various settings, the biases and personalities of investigators, and the pressures and ideologies to which they may be subject, we may learn not only how to account for and reduce the kinds of discrepancies found in the present study, but also how to develop more powerful means for the prediction and control of human behavior.

While the findings of this study have implications beyond the realm of Rorschach validity research, their importance for this particular area is considerable. Together with demonstrations of examiner influences on test protocols, they make a complete examination of the research procedures and behaviors of Rorschach investigators imperative if any sense is to be made out of the swelling literature dealing with the test.

SUMMARY

In an attempt to determine the extent to which sociological variables may be important in contributing to the kind of research done and probabilities of certain outcomes, 168 Rorschach test validity studies appearing in four journals during a five-year period were culled according to certain criteria. These studies were classified on three dichotomized variables: each study came from an academic or nonacademic setting, was primarily con-

cerned with construct or criterion validity, and was interpreted by its author as yielding results which were either favorable or unfavorable to the Rorschach. Four hypotheses were formulated and tested regarding the relationships between these variables; each was substantially supported. The implications of these findings for the importance of the social psychology of research as an area of investigation were discussed with particular emphasis on the problems involved in the evaluation of current Rorschach test validity research.

REFERENCES

1. ALDEN, PRISCILLA, & BENTON, A. L. Relationship of sex of examiner to incidence of Rorschach responses with sexual content. *J. proj. Tech.*, 1951, **15**, 231-234.
2. CRONBACH, L. J. Statistical methods applied to Rorschach scores: A review. *Psychol. Bull.*, 1949, **46**, 393-429.
3. CRONBACH, L. J., & MEEHL, P. E. Construct validity in psychological tests. *Psychol. Bull.*, 1955, **52**, 281-302.
4. CURTIS, H. S., & WOLF, ELIZABETH B. The influence of the examiner on the production of sex responses on the Rorschach. *Amer. Psychologist*, 1951, **6**, 345-346. (Abstract)
5. GIBBY, R. G. Examiner influence on the Rorschach inquiry. *J. consult. Psychol.*, 1952, **16**, 449-455.
6. GIBBY, R. G., MILLER, D. R., & WALKER, E. L. The examiner's influence on the Rorschach protocol. *J. consult. Psychol.*, 1953, **17**, 425-428.
7. HOLTZMAN, W. H. The examiner as a variable in the Draw-a-Person Test. *J. consult. Psychol.*, 1952, **16**, 145-148.
8. LEVY, L. H. A note on research methodology used in testing for examiner influence in clinical test performance. *J. consult. Psychol.*, 1956, **20**, 286.
9. MANNHEIM, K. *Ideology and utopia: An introduction to the sociology of knowledge*. New York: Harcourt, Brace, 1936.
10. MILLER, D. R. Prediction of behavior by means of the Rorschach test. *J. abnorm. soc. Psychol.*, 1953, **48**, 367-375.
11. NORTON, H. W. Calculation of chi square for complex contingency tables. *J. Amer. stat. Assn.*, 1945, **40**, 251-258.
12. SANDERS, R., & CLEVELAND, S. E. The relationship between certain examiner personality variables and subject's Rorschach scores. *J. proj. Tech.*, 1953, **17**, 34-50.

Received September 23, 1957.

MULTIVARIABLE ANALYSIS OF THE CONCEPTUAL BEHAVIOR OF SCHIZOPHRENIC AND BRAIN-DAMAGED PATIENTS

DONALD B. LEVENTHAL¹

Topeka State Hospital

LAURENCE S. MCGAUGHRAN²

University of Houston

AND LOUIS J. MORAN

University of Texas

MOST studies of differences between schizophrenic and organically based disorders in thinking have employed either a form of analysis derived from the Goldstein-Scheerer (1941) dichotomy of abstract-concrete "attitudes" or one which assesses group differences in terms of "intellectual deterioration." Since both of these conceptions imply that thinking occurs as a function of a single "ability" or faculty—"abstract attitude" or "intellect"—they can provide reliable distinctions between schizophrenics and the brain-injured only in terms of differences in degree of loss of this ability. Thus, when differences are found with these forms of analysis, the organic group usually (but not always) shows greater concreteness (nonabstractness) or greater ineptness of performance in "intellectual" tasks. In addition to objections concerning limitations upon the interpretation of such findings, there is the added difficulty cited by Yates (1954) that the effects of possibly important control factors (e.g., age, intelligence, education, sex) usually have been ignored. Consequently, one cannot be sure of the sources to which such group deficits in a unitary ability can be attributed.

There have been a few investigators (Cameron, 1938; Hunt & Cofer, 1944; Hutson & Shakow, 1949; Lidz, Gay, & Tietze, 1942; Rashkis, 1947) who have attempted distinctions between schizophrenic and brain-damaged conceptualization on bases other than the existence of greater or lesser impairment of abstract or of intellectual ability. More recently, studies by McGaughran (1954), McGaughran and Moran (1956; 1957), Moran,

McGaughran, and Leventhal (1957), and by Grassi (1947; 1953) have introduced two schemas permitting a multivariable analysis of conceptual behavior. Each of these schemas has been employed to assess group differences in conceptualization between schizophrenic and brain-damaged patients.

One schema employs, at the present, two conceptual variables, "amount of social agreement" and "order of conceptual classification." In one study (McGaughran & Moran, 1956), it was successfully predicted that the major difference between schizophrenic and non-psychiatric groups is the amount of social agreement or communality characterizing the concepts they employ. In another study (McGaughran & Moran, 1957), order of conceptual classification showed the greater difference between schizophrenic and brain-damaged Ss, while the scores on this measure for the nonpsychiatric group fell between those of the two deviant groups. It was concluded that: "The differences in conceptual behavior between schizophrenic and brain-damaged patients suggest that the two groups should not be represented as evidencing the same type of conceptual disorder" (McGaughran & Moran, 1957, p. 48).

Grassi (1953) reported a high degree of success in differentiating conceptual performance among normal, schizophrenic, and brain-damaged groups. His schema may also be interpreted as consisting of two conceptual variables: "abstract-concrete" and "complex-simple." It is only fair to state that Grassi apparently has not considered his method in this way. For him (1953), the terms "simple" and "complex" only convey two levels of difficulty in tasks involving either abstract or concrete performance. Further, the single score used in the Grassi test is ob-

¹ Formerly at the University of Houston.

² Employed part time at the Houston Veterans Administration Hospital, where part of the data analysis of this study was conducted.

tained by summing the subscores for performance at both levels of difficulty for both concrete and abstract tasks. This usage of a single score is clearly consistent with the Goldstein-Scheerer (1941) formulation of the presence-absence (or loss) of abstract ability. Indeed, Grassi (1953) states that "the theory and rationale set forth by Goldstein and Gelb is that upon which the Grassi test is predicated" (p. 8). He considers his test simply as a more sensitive index of "impairment in both the concrete and abstract spheres" (p. 13). Considering some of Grassi's quantitative results and some of his interpretations of his data, however, it may be rewarding to examine his instrument from the viewpoint of an explicit multivariable form of analysis.

The primary purpose of this study was to cross-validate the research findings of McGaughan and Moran (1957) and Grassi (1953) of differences in conceptual behavior between schizophrenic and brain-injured groups. Three other related purposes were (a) to extend the "range of convenience" of the McGaughan and Moran schema by applying it under different operational conditions, (b) to assess the predictive efficiency of both systems of conceptual analysis, and (c) to investigate the feasibility of considering the Grassi schema as one providing the opportunity for a multivariable analysis of conceptual behavior.

METHOD

Subjects

Ss were 60 male patients at Topeka State Hospital, Topeka, Kansas. Psychiatric diagnoses served as the criteria for selection. Thirty patients diagnosed as schizophrenic reaction, paranoid type, who had been in the hospital for at least two years made up the schizophrenic group. Thirty patients diagnosed as showing a chronic brain syndrome (without psychosis) associated with cerebral arteriosclerosis, senile brain disease, alcoholism, brain trauma, or central nervous system syphilis composed the brain-damaged group. There were no significant differences between the schizophrenic and brain-syndrome groups in age, education, or estimated intelligence level.³

³ Upon request, mimeographed copies of a table of means and standard deviations of the distributions for these variables will be provided by the writers.

Procedure

One examiner administered individually to all Ss the following tests of concept formation: (a) an object-sorting task, (b) a similarities task, and (c) the Grassi Block-Substitution Test. Before scoring was undertaken, the identity of the individual test protocols was concealed by means of coding. Each separate item in the object-sorting and similarities tasks was scored for all Ss; then the next item was scored in a similar way, and so on. Interscorer agreement on the object-sorting and similarities tasks was 90% and 86%, respectively; because scores on the Grassi test are based simply upon performance accuracy and time, no assessment of scoring reliability seemed necessary.

Object-sorting task. Introduced by Goldstein and Gelb and modified by a number of others, the object-sorting task requires 19 different sortings of a group of 32 objects familiar in everyday experience. The task materials and administration corresponded to those described by Rapaport (1946) and previously employed by McGaughan and Moran. The 19 responses by each S were scored according to the schema described previously (McGaughan & Moran, 1956; 1957). In this schema,⁴ extent of social agreement (or "publicness-privateness") and order of conceptual classification (or "openness-closedness") are intersected to form four scoring quadrants termed closed-public, open-public, open-private, and closed-private. These quadrants or "conceptual areas" are the basic scoring categories; each response is classified as occurring within one of the four.

The determination of the appropriate quadrant or area for a concept requires a cross-classification in which the effects of both variables can be taken into account with one notation. By summing the individual S's scores in adjacent quadrants, two additional measures, "total public vs. private" and "total open vs. closed," can be derived. These measures provide a notation of the effect of either variable, independent of the other. One additional measure, an "autistic index," is derived by subtracting the sum of closed-public concepts from the sum of concepts scored as open-private.

Similarities task. The items on the similarities task consisted of groups of two to four words which could be "categorized" under a single concept. Seventeen such groups of words were presented orally one at a time to the Ss, who were instructed to "tell what the words have in common—that is, in what way they are alike."

The scoring system developed (Moran, 1953; Moran et al., 1957) to classify each response is based upon the same multivariable form of analysis employed in the object-sorting task. Thus, although the two tasks differ considerably in a number of respects, the same

⁴ A revised set of subarea categories and supplementary instructions for scoring conceptual areas as used in this study has been deposited with the American Documentation Institute. Order Document No. 5067 from the ADI Auxiliary Publications Project, Photo Duplication Service, Library of Congress, Washington 25, D.C., remitting in advance \$1.75 for microfilm or \$2.50 for photocopies. Make check payable to Chief, Photoduplication, Library of Congress.

seven measures used for object-sorting may be applied to the similarities task.

Grassi Block-Substitution Test. Grassi's (1953) test consists of a set of 24 cubes of identical size, painted different colors on the different sides. Five large blocks were made out of 20 of the regular cubes; each block was composed of four cubes glued together to form a specific pattern. The four remaining cubes were to be put together by *S* in four different ways for each of five designs, using a model presented by the examiner as a referent. The four types of tasks required copying the block model on the top, on the top but in a different color, on the top and sides in the same color, and on the top and sides but in a different color. Levels of conceptual performance on these four types of tasks were characterized by Grassi as simple-concrete, simple-abstract, complex-concrete, and complex-abstract, respectively.

In Grassi's study, *S*'s performance on each of the 20 tasks was scored on the basis of time and accuracy credits, but he employed only a single total performance score⁵ in arriving at an estimate of group differences. He also reported, however, the number of complete failures for each experimental group on each of the 20 tasks. To investigate the feasibility of adapting the Grassi schema to a multivariable form of analysis, the test data in the present study were scored not only in terms of total performance and number of complete failures, but also in general accordance with the multivariable schema employed in the analysis of responses in the object-sorting and similarities tasks. Thus, group differences in performance scores were assessed for each of the four "levels of performance" or scoring "quadrants" in the Grassi schema.

Hypotheses

Object-sorting and similarities tasks. The initial investigation by McGaughran and Moran (1957) dealt with a specific population (VA hospital patients) and a specific task (object-sorting). The present study is intended not only to check the validity of these findings in the usual sense of broadening the reference population, but also in terms of construct validity—i.e., testing a designated schema of multivariable analysis of conceptual behavior under varying operational conditions. Thus, since parallel systems of scoring were developed for the object-sorting and similarities tasks, the same hypotheses are presented for both sets of tasks.

The general hypothesis of differences between brain-damaged and schizophrenic groups in both object-sorting and similarities tasks is that the brain-damaged group has a greater deficit in higher order concepts (loss of "openness").

The specific hypotheses stated in terms of the operations of both tasks in the study are as follows: In comparison with the schizophrenic group, the brain-damaged group employs (a) significantly more closed concepts; (b) significantly fewer open-private, (c) and significantly more closed-public and (d) closed-private concepts; and (e) scores significantly lower on an autistic index consisting of open-private minus closed-public concepts.

⁵ Complete scoring instructions may be found in Grassi's (1953) test manual.

Grassi Block-Substitution Test. In accordance with Grassi's (1953) findings, the specific hypotheses, stated in terms of the operations of his study, are as follows. In comparison with the schizophrenic group, the brain-damaged group (a) scores significantly lower on total score and (b) demonstrates significantly more complete failures on all four levels of performance but with more extreme group differences in performance occurring in the simple-concrete and complex-concrete tasks.

To test the feasibility of adapting the Grassi schema to a multivariable form of analysis, it was further hypothesized that brain-damaged patients score significantly lower on the (a) simple-concrete and (b) complex-concrete levels of performance.

RESULTS

To provide a statistical control over the effects of intelligence, group differences on each measure were calculated by analyses of covariance except for counts of complete failures on the Grassi test; for these, analyses of differences in percentage were employed. The tables are arranged so that, when appropriate, the results in the present investigation may be compared directly with those obtained in the initial studies.

The results for the object-sorting measures for both the original and the present study are shown in Table 1. In the present study, all five measures involved in the hypotheses show significant mean differences in the direction predicted.

The results for the similarities task are shown in Table 2. In this task, all five of the measures involved in the hypotheses show group mean differences in the direction predicted, and four of these (excluding closed-private) are significant at the .01 level or better.

On the Grassi test, the group difference on total score was in the predicted direction and significant at the .05 level. While it is not possible to compute the significance of group differences in the Grassi study from the data provided, a comparison⁶ of differences in mean, range, and group overlap between the two studies indicates quite clearly that the original groups differed at a considerably higher level of significance than the groups in the present study.

The distribution of complete failures for each of the four performance levels on the Grassi test for both studies is shown in Table 3. It can be seen that in the Grassi study the

⁶ Upon request, mimeographed copies of a table showing these comparisons will be provided by the writers.

TABLE 1

COMPARISON OF GROUP DIFFERENCES OBTAINED IN THE MCGAUGHRAN AND MORAN STUDY AND IN THIS STUDY IN CONCEPTUAL AREA SCORES ON AN OBJECT SORTING TASK

Measure	McGaughran and Moran Study						Present Study					
	Schizophrenic (<i>N</i> = 37)		Brain-Damaged (<i>N</i> = 34)		<i>F</i> and <i>p</i> Value		Schizophrenic (<i>N</i> = 30)		Brain-Damaged (<i>N</i> = 30)		<i>F</i> and <i>p</i> Value	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	Adjusted ^a <i>F</i> Ratio	<i>p</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	Adjusted ^a <i>F</i> Ratio	<i>p</i>
Public	11.00	4.70	11.27	3.67	0.82	—	8.50	4.89	12.10	5.03	13.49	< .01
Closed	6.05	3.35	11.18	3.29	25.34	< .001	5.03	3.77	9.87	3.19	27.34	< .001
Closed-Public	3.78	2.62	6.77	3.46	14.71	< .001	4.00	3.37	6.90	3.09	11.59	< .01
Closed-Private	2.27	2.11	4.41	2.93	3.78	< .06	1.03	1.11	2.97	1.83	22.84	< .001
Open-Public	7.22	3.98	4.50	2.74	2.43	—	4.50	4.28	5.20	3.51	1.47	—
Open-Private	5.73	3.95	3.32	2.48	10.11	< .01	9.47	4.99	3.93	2.84	32.25	< .001
Autistic ^b index	11.95	5.95	6.55	5.13	22.99	< .001	15.47	7.40	7.03	4.78	27.28	< .001

^a Analysis of covariance was used to control for the effects of intelligence on task performance.^b Open-private minus closed-public, plus 10.

TABLE 2

SIGNIFICANCE OF DIFFERENCES IN CONCEPTUAL AREA SCORES ON A SIMILARITIES TASK BETWEEN SCHIZOPHRENIC AND BRAIN-DAMAGED GROUPS

Measure	Schizophrenic (<i>N</i> = 30)		Brain-Damaged (<i>N</i> = 30)		Adjusted ^a <i>F</i> Ratio	<i>p</i>	<i>r</i>	<i>r_p</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>				
Public	8.57	4.34	12.03	3.14	13.49	< .01	.24	—
Closed	6.03	3.43	10.63	3.21	25.68	< .001	-.41	< .01
Closed-Public	4.40	3.28	8.63	3.16	25.01	< .001	-.35	< .01
Closed-Private	1.63	1.84	2.00	1.59	.45	—	-.12	—
Open-Public	4.17	3.05	3.40	3.19	.30	—	.58	< .001
Open-Private	6.80	3.60	2.97	2.21	26.21	< .001	-.23	—
Autistic Index ^b	12.40	6.03	4.33	4.62	32.05	< .001	.09	—

^a Analysis of covariance was used to control for the effects of intelligence on task performance.^b Open-private minus closed-public plus 10.

TABLE 3

COMPARISON OF GROUP DIFFERENCES OBTAINED IN THE GRASSI STUDY AND IN THE PRESENT STUDY ON PERCENTAGE OF ACTUAL FAILURES AT EACH LEVEL OF PERFORMANCE ON THE GRASSI BLOCK SUBSTITUTION TEST

Level of Performance	Grassi Study ^a				Present Study			
	Schizo- phrenic (<i>N</i> = 86) ^b	Brain- Damaged (<i>N</i> = 102) ^b	CR	<i>p</i>	Schizo- phrenic (<i>N</i> = 30) ^b	Brain- Damaged (<i>N</i> = 30) ^b	CR	<i>p</i>
Simple-Concrete	0.0	10.2	7.28	< .001	1.3	19.3	5.51	< .001
Simple-Abstract	5.1	43.9	13.86	< .001	30.0	43.3	2.38	< .05
Complex-Concrete	1.4	47.2	16.36	< .001	20.0	42.0	4.15	< .001
Complex-Abstract	40.9	90.6	16.56	< .001	68.0	82.0	2.80	< .01

^a Figures derived from Grassi's (1953, p. 65) table of number of failures on each of 5 tasks on each level of performance.^b *N*s shown are for number of *S*s; the total number of tasks failed and passed at each level of performance equals number of *S*s multiplied by 5.

brain-damaged group exceeded the schizophrenics in number of failures at all levels of performance beyond the .001 level of significance. On the other hand, in the present study, the *p* values for group differences in number of failures for the two performance

levels involving abstraction drop to .05 and .01. This, of course, means that there was less group distinctiveness in impairment or "intellectual deterioration" when abstractive ability was involved.

Scores on the Grassi test for each of the

TABLE 4
SIGNIFICANCE OF DIFFERENCES IN LEVELS OF PERFORMANCE SCORES ON THE GRASSI BLOCK
SUBSTITUTION TEST BETWEEN SCHIZOPHRENIC AND BRAIN-DAMAGED GROUPS

Level of Performance	Schizophrenic (N = 30)		Brain-Damaged (N = 30)		Adjusted F Ratio ^a	p	r	r _p
	M	SD	M	SD				
Simple-Concrete	4.93	0.08	4.37	1.38	4.68	< .05	.42	< .01
Simple-Abstract	3.50	1.71	2.83	1.69	1.81	—	.54	< .001
Complex-Concrete	4.00	1.63	2.90	2.14	4.49	< .05	.19	—
Complex-Abstract	1.60	1.69	0.90	1.35	2.40	—	.51	< .001

^a Analysis of covariance was used to control for the effects of intelligence on test performance.

four levels of performance are shown in Table 4. The results indicate that, as predicted, the brain-damaged group shows significantly ($p = .05$) more impairment than the schizophrenics on the simple-concrete and complex-concrete levels of performance, but the groups do not differ significantly on the performance levels involving the abstractive process.

In a final analysis of the Grassi test data, rank-order correlations ($N = 60$) were computed between all possible combinations of score distributions for the four performance levels. The values for the correlations were all positive and ranged from .57 (for complex-concrete vs. complex-abstract) to .72 (for simple-concrete vs. complex-abstract). All correlation values were significant at the .001 level. A corresponding analysis of the amount of association between the "publicness-private-ness" and "openness-closedness" variables indicated that distributions of scores obtained in these measures were not significantly related in either the object-sorting or similarities tasks.

DISCUSSION

Efficiency of Individual Prediction

The total group mean was used as a cutting point for the major predictor variables on each of the three tasks; for the object-sorting and similarities tasks, the major predictor was the sum of "closed" concepts and, for the Grassi test, it was the total score. On the object-sorting task, the number of correct predictions for the schizophrenic group was 24 out of 30 (80%); for the brain-damaged group, it was 25 out of 30 (83%). The similarities task data yielded correct predictions for 23 out of 30 Ss (77%) in each group. The number of correct predictions obtained⁷ with the Grassi test was

⁷ Had the cutting point suggested by Grassi (1953, p. 30) been used, the percentage of correct predictions

19 out of 30 (63%) for the schizophrenics and 14 out of 30 (47%) for the brain-syndrome group.

When one considers the amount of selectivity employed in this and in similar studies to form relatively "pure" experimental groups, (e.g., selecting brain-damaged cases without psychosis), it seems reasonable to conclude that none of the three tasks in its present form appears to be a very effective clinical instrument for independent diagnosis of brain damage. However, it seems likely that predictive validity in this area might well be increased through the development of a more sensitive set of task or scoring characteristics within the present schematic contexts.

It should be stated that Grassi quite clearly emphasized that scores on his test should never be used apart from qualitative behavioral signs in making clinical judgments. Nevertheless, in addition to Yate's (1954) criticism of this stand, it may be added that most of the behavioral signs (e.g., perseveration, requests for reassurance) suggested by Grassi are in no way specific to his test and, consequently, would not enter into a judgment of its diagnostic effectiveness.

Cross-Validation of the Two Schemas

"Conceptual area" schema. An analysis of the object-sorting data in Table 1 indicates that the findings of group differences in the two studies were much the same, except that there was a much stronger inclination for the schizophrenic group in the present study to use open-private sortings at the expense of lowered scores on the open-public measure. This difference can possibly be accounted for in terms of a corresponding difference in sampling. The significantly ($p = .001$) older would have been 33% for the schizophrenic group and 67% for the group with brain-damage.

schizophrenics of the present sample had, for the most part, been hospitalized for a number of years as contrasted with the younger group of VA patients used in the original study. The greater degree of chronicity for the present group, then, may be associated with a greater tendency to employ open-private sortings, a type of conceptualization which presumably would increase with increased social withdrawal.

Closely equivalent results were obtained in the present study with the object-sorting and similarities tasks. Since the same multivariable schema was employed to develop corresponding systems of scoring under rather widely different task conditions, the conceptual variables of amount of social agreement and order of conceptual classification do not seem bound to any one set of operations. The major difference between schizophrenic and brain-damaged groups in their performance on both conceptual tasks was in terms of order of conceptual classification. As in the original McGaughran and Moran study, these differences are characterized as tendencies toward "under-abstraction" in the brain-damaged group and "over-abstraction" in the schizophrenics.

Grassi "single total score" schema. When the variables of age, sex, intelligence, and education were controlled, the brain-damaged group still showed a significantly greater "intellectual deterioration" (Grassi, 1953, p. 30). However, the theoretical implications of such a finding seem rather limited. In the absence of any further denotation, the determination of greater or lesser "impairment in both the concrete and abstract spheres" (p. 13), conveys nothing except more or less accuracy or speed in a series of tasks which, since the scores are simply additive, presumably vary only in level of difficulty.

As for the practical implications of the results for the total score measure, there are rather striking differences between the findings in the two studies. Although it is not entirely clear, it seems likely that these discrepancies may be attributed to differences in controls and in sampling. Concerning controls, Grassi states, "... it was felt that the test would be at fullest efficiency if employed with subjects above the defective level. No limitations as to age, sex, or any other factor were found necessary" (1953, p. 60). Since age, sex, education, and

intelligence were controlled in the present study, a considerable amount of the discrepancy between the two studies may be attributed to differences in the effects of these factors.

Concerning sampling, the greatest part of the discrepancy in the findings of the two studies consists of differences between the schizophrenic groups. For example, the schizophrenics in the present study scored significantly more failures at every level of performance than the corresponding group in the Grassi study. Grassi reported that he chose equal numbers of "deteriorated" and "non-deteriorated" schizophrenics "without regard to sub-classification" (1953, p. 62), and concluded at the end of his study that "non-deteriorated schizophrenics reveal no significant degree of impairment" (p. 66). However, he offered no criteria to distinguish between deterioration and nondeterioration, and this distinction is maintained nowhere in the actual analysis of his data nor in his setting of cutting points. Thus, it is not possible to make a definitive analysis of differences in amounts of "deterioration" in the two schizophrenic samples, except to state that the group in the present study presumably evidenced more of it.

Grassi Schema as a Multivariable Form of Analysis

In terms of the frequency of complete failures in the Grassi test, there was less group distinctiveness between schizophrenics and the brain-damaged on the two levels of performance in which abstractive ability was involved. This could conceivably be interpreted to mean that the simple-complex and concrete-abstract measures represent somewhat different variables of conceptualization and that, consequently, the Grassi schema could be adapted to provide a system of multivariable analysis of conceptual behavior.

There is one rather serious limitation to such an interpretation. To coincide with Grassi's original procedure, the analysis of failures in the present study was in terms of total number, regardless of the number of Ss involved. When the present data were reanalyzed in terms of number of Ss showing failures at each of the performance levels, none of the group differences proved to be significant. However, this difficulty was resolved to some extent

(although variances were not homogeneous) by turning to an analysis of group differences in mean scores at each of the four performance levels. Here again the brain-damaged group showed significantly greater impairment than the schizophrenics in tasks involving simple perceptual fidelity with varying degrees of complexity, but the group difference in impairment disappeared in the two tasks requiring an abstractive reduction of visual images.

While there seems to be some basis for inferring that the simple-complex and concrete-abstract measures in the Grassi test may represent somewhat different sources of variation in conceptual behavior, the degree of intercorrelation among the scores at all levels of performance indicates that the measures are far from independent, at least as they are presently employed. Greater separateness or independence of these variables may be achieved by a systematic alteration of task requirements.

SUMMARY

The primary purposes of this study were to attempt to cross-validate previous findings of two schemas for analyzing differences in conceptualization between schizophrenic and brain-damaged groups, and to investigate the feasibility of adapting one of these (the Grassi schema) into a multivariable system of analysis corresponding in form (multivariable interaction) but not in content with the other ("conceptual area") schema, used previously by the present investigators.

Object-sorting, similarities, and block-substitution tasks were employed. The "conceptual area" schema correctly predicted hypothesized group mean differences ($p = .01$ or $.001$) in all five of the measures in object-sorting and in four out of five in the similarities task. The Grassi schema successfully predicted group mean differences ($p = .05$) in total score on the block-substitution task.

Further analysis of the Grassi test data indicated that the brain-damaged group showed greater impairment in tasks involving simple perceptual fidelity but not in tasks

requiring an abstractive reduction of visual images. It was suggested that greater independence of these conceptual variables may be achieved by experimentally manipulating task requirements.

REFERENCES

- CAMERON, N. Reasoning, regression, and communication in schizophrenics. *Psychol. Monogr.*, 1938, **50**, No. 1 (Whole No. 221).
- GOLDSTEIN, K., & SCHEERER, M. Abstract and concrete behavior. *Psychol. Monogr.*, 1941, **53**, No. 2 (Whole No. 239).
- GRASSI, J. R. The Fairfield Block Substitution Test for measuring intellectual impairment. *Psychiat. Quart.*, 1947, **21**, 474-489.
- GRASSI, J. R. *The Grassi Block Substitution Test for measuring organic brain pathology*. Springfield, Ill.: Charles C Thomas, 1953.
- HUNT, J. McV., & COFER, C. N. Psychological deficit. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. 2. New York: Ronald, 1944. Pp. 971-1032.
- HUTSON, P. E., & SHAKOW, D. Learning capacity in schizophrenia—with special reference to the concept of deterioration. *Amer. J. Psychiat.*, 1949, **105**, 881-888.
- LIDZ, T., GAY, J. R., & TIETZE, C. Intelligence in cerebral deficit states and schizophrenia measured by the Kohs Block Test. *Arch. Neurol. Psychiat.*, 1942, **48**, 568-582.
- MCGAUGHRAN, L. S. Predicting language behavior from object-sorting. *J. abnorm. soc. Psychol.*, 1954, **49**, 183-195.
- MCGAUGHRAN, L. S., & MORAN, L. J. "Conceptual level" vs. "conceptual area" analysis of object-sorting behavior of schizophrenic and nonpsychiatric groups. *J. abnorm. soc. Psychol.*, 1956, **52**, 43-58.
- MCGAUGHRAN, L. S., & MORAN, L. J. Differences between schizophrenic and brain-damaged groups in conceptual aspects of object-sorting. *J. abnorm. soc. Psychol.*, 1957, **54**, 44-49.
- MORAN, L. J. Vocabulary knowledge and usage among normal and schizophrenic subjects. *Psychol. Monogr.*, 1953, **67**, No. 20 (Whole No. 370).
- MORAN, L. J., MCGAUGHRAN, L. S., & LEVENTHAL, D. Intra-individual consistency in conceptualization. *J. abnorm. soc. Psychol.*, 1957, **55**, 194-196.
- RAPAPORT, D., GILL, M., & SCHAFER, R. *Diagnostic psychological testing*. Vol. II. Chicago: Year Book Publishers, 1946.
- RASHKIS, A. R. Three types of thinking disorder. *J. nerv. ment. Dis.*, 1947, **106**, 650-670.
- YATES, A. J. The validity of some psychological tests of brain-damage. *Psychol. Bull.*, 1954, **51**, 359-379.

Received September 23, 1957.

A GENERAL FORMULA FOR THE QUANTITATIVE TREATMENT OF HUMAN MOTIVATION

WALTER TOMAN

Brandeis University

MAN always is motivated, but also is satisfying motives at the same time. While preparing food in order to eat, he is satisfying the motive to handle and manipulate things. Yet while eating, he still is motivated to get food on his plate, to cut his meat, or to chew it. Satisfactions of motives, however, can be defined, among other ways, by the decrease of motivation that they bring about. And the decrease of motivation can, in turn, be determined by means of subsequent satisfactions. Other things being equal, the satisfaction value of a meal, i.e., its capacity to reduce a person's hunger, can be inferred from the time at which he chooses to have his next meal. What holds for eating holds for any recurrent motive that can be identified, whether it is something as primitive and "physiological" as moving or even breathing or something as complicated as the motive to be in company or to write a paper. All of a person's motives are forever on the increase from the moment they were last satisfied until they are satisfied again. On the other hand, satisfactions (of various degrees) reduce motive intensities (by corresponding amounts). Some motives are always more urgent than others, but when they are satisfied, they make room for others. They fall back toward the rear of the race.

This race changes its complexity as the individual develops. New motives are being formed all the time. The only thing the newborn can do effectively and somewhat by intention is sucking. At the age of one year, however, he can suck, bite, grasp, hold, and move various things, he can crawl, sit, stand, etc. And the intervals between successive satisfactions of motives tend to increase and to become more variable. The newborn can keep awake for an hour at a time, and the adult for 16 hours. The newborn falls asleep under almost any circumstances after one-and-a-half hours of waking. The adult can, if necessary, go without sleep even after 36-hours of waking. Similar conditions could be shown to prevail

with other recurrent motives, although variations of opportunity to satisfy motives as well as the fact that motives can do something for each other to various degrees (i.e., substitutability) complicate matters. Anyway, the race increases in complexity with a person's psychological development. It slows up, and more things happen simultaneously.

A spectator trying to keep track of this race of motives within a person is hopelessly lost unless he focuses on only a few at a time. To do so throughout a person's course of development is, of course, out of the question. What may be possible, however, is to trace a sample of a person's motives over a certain stretch of time and estimate the kind of race that is being run. If we can find motives that are being satisfied recurrently by everyone and that are relatively uncomplicated by differences of opportunity and by substitute motives, or if we can interpolate for differences of opportunity and include at least the most relevant substitute motives, we may be able to infer something very general and important about a person in comparative, not absolute, terms: the level of his motivational development.

For this purpose, the following general formula is offered:

$$\sum_{i=1}^n \epsilon_i = C$$

In words: the sum of intensity increments of motives that can be distinguished within a given person at a given time of development is equal to a constant. This assumption is claimed to be the simplest of all possible holistic assumptions about a person's motivation.

In the formula, ϵ_1 stands for intensity increment of Motive 1, say sleeping; ϵ_2 for that of Motive 2, say eating; ϵ_3 for that of talking, etc. These increments can be estimated from the average intervals (i) between successive satisfactions of each motive by taking their reciprocals ($\epsilon = \frac{1}{i}$). If a person's average

interval between successive occasions of sleep is 16 hours, the intensity increment per hour is $\frac{1}{16}$; if he eats every 4 hours, the intensity increment per hour is $\frac{1}{4}$; etc. They can also be estimated from the differences between maximal and minimal intervals (d) between successive satisfactions ($\epsilon = \frac{1}{d}$). If a person's

minimal interval between successive occasions of sleep has been 6 hours, say during the past month, and his maximal interval 36 hours, his range (d) would be 30 hours, and his intensity increment $\frac{1}{30}$ per hour. These two types of estimates cannot be mixed within one sample. While the interval between successive satisfactions is the time from the end of one to the onset of the subsequent satisfaction of a motive, it may often suffice to measure or estimate intervals from onset to onset of successive satisfactions.

Letter n stands for the number of motives that can be identified within a person. Motives 1, 2, 3, . . . $m-2$, $m-1$, are those that have already been satisfied, motives in the state of recurrence, such as sleeping, eating, talking, etc. Motives m , $m+1$, $m+2$, . . . $n-2$, $n-1$, n , are motives that have not yet been satisfied, motives in the process of formation, such as graduating from high school or college, owning a house, marrying a certain person, becoming competent in electronics, etc. They represent the "free energy portion" of a person.

The C can be taken as a constant for a given individual throughout his development. This is, if nothing else, the simplest assumption about C and at least worth a try. C can perhaps be understood as the over-all motivational energy with which a person operates throughout his life.

The only additional assumption we need make about motivational development is that the number of distinguishable motives (n) increases with development. Therefore, the intensity increments with which any given recurrent motive grows until it is once more satisfied must, on the whole, decrease with development. Empirical evidence has already been indicated.

The motive to sleep, e.g., takes an hour with the newborn to reach a given intensity, that at which satisfaction usually occurs, whereas it takes 16 hours with the adult. Hence the intensity increment per hour with which

the adult's motive to sleep grows in intensity is $\frac{1}{16}$ of that of the newborn.

Since n becomes impracticably large even early in a person's development, since new motives and new substitute motives are continually being formed, since there are even substitutions by force (due to psychological losses, e.g., weaning), and since opportunity to satisfy motives is such a complicated and variable thing, this formula can be applied in practice only with discretion by means of appropriate sampling and in approximations. That, however, is possible. This paper presents nomothetic evidence to test all aspects of the formula with subjects or groups of subjects about whose developmental differences there can be little doubt.

EXPERIMENTAL EVIDENCE

Decreasing Increments

Let us see whether single motives reflect the developmental trend of decreasing increments (or of increasing average intervals between successive satisfactions and increasing ranges of intervals). Let us take sleeping, which is likely to be highly uncomplicated by differences of opportunity. A person can usually impose the amount of sleep he needs on the circumstances of his life. In fact, he would not live for long if he could not. Let us take eating (meals and snacks) which is a little more opportunity-dependent. Let us take the motive to have refreshments (such as Coca-Cola, milk shake, ice cream, coffee, etc., consumed in cafeterias, drugstores, and the like, but not together with a meal), and the motive to visit movie theaters, both of which are obviously highly determined by opportunity. Let us study these motives in high school sophomores, high school seniors and college juniors, of both sexes, ethnically and socioeconomically matched, 15, 17, and 20 years old on the average, and ask them to rate their average, their largest, and their smallest interval between successive periods of sleep for the past two months, between successive meals plus snacks (ignoring the night interval) for the past two months, and between successive visits to movie theaters and to refreshment parlors for the past year. The subjects' estimates of average intervals will be used directly, whereas their estimates of largest and smallest intervals will be transformed into the range d

TABLE 1
AVERAGE INTERVALS BETWEEN SUCCESSIVE
SATISFACTIONS OF MOTIVE
(Means of estimated averages)

Subjects (age groups)	N	Sleeping (hours)	Eating (hours)	Visiting Movie Theaters (days)	Having Refresh- ments (days)
15 years	29	15.3	3.8	29.5	4.8
17 years	19	15.7	4.1	26.7	5.3
20 years	32	16.0	4.2	18.9	8.6

TABLE 2
RANGES OF INTERVALS BETWEEN SUCCESSIVE
SATISFACTIONS OF MOTIVE
(Mean differences between estimated largest and
estimated smallest interval)

Subjects (age groups)	N	Sleeping (hours)	Eating (hours)	Visiting Movie Theaters (days)	Having Refresh- ments (days)
15 years	29	6.9	6.2	76.9	26.9
17 years	19	10.7	7.1	76.4	17.1
20 years	32	13.9	7.6	46.7	32.1

(difference between the maximal and the minimal interval between successive satisfactions).

Tables 1 and 2 show the results. Data for sleeping and eating confirm significantly the predicted developmental trend (all t tests of difference are significant at least on the .05 level), whereas visiting movie theaters and having refreshments behave more irregularly, in ways perhaps attributable to differences in opportunity.

Let us see whether a small sample of motives also reflects the developmental trend of decreasing increments, perhaps more adequately than a single arbitrarily chosen motive. The sum of intensity increments of the motives sleeping, eating, visiting movie theaters, and having refreshments should decrease with age. Computing intensity increments from the reciprocals of the average intervals between successive satisfactions (from Table 1) yields the following for the 15-year-old group:

$$\sum_{i=1}^4 \epsilon_i = \frac{1}{15.31} + \frac{1}{3.77} + \frac{1}{29.45} \times \frac{1}{24} + \frac{1}{4.78} \times \frac{1}{24} = 0.34.$$

Similarly, sums of 0.31 and 0.30 were found for the 17- and 20-year-olds respectively. Computing intensity increments from the

TABLE 3
SUMS OF INTENSITY INCREMENTS FOR FOUR MOTIVES
(Indices based on data in Tables 1 and 2)

Subjects (age groups)	N	$\Sigma \epsilon_{(i)}^a$	$\Sigma \epsilon_{(d)}^b$	$\Sigma \epsilon_{(d)} / \Sigma \epsilon_{(i)}$
15 years	29	0.34	0.31	0.92
17 years	19	0.31	0.24	0.76
20 years	32	0.30	0.21	0.68

^a Estimate based on summed reciprocals of average intervals between satisfactions, from Table 1.

^b Estimate based on summed reciprocals of range of intervals between successive satisfactions, from Table 2.

reciprocals of ranges of intervals (given in Table 2) shows the same thing: the sum of intensity increments decreases with development (see Table 3). A parallel analysis based on the averaging of measures of intensity increment computed for each individual separately shows the same trend, significant at the .05 level.

One can assume on various grounds that the variability, in our case the range, of intervals increases more rapidly with development or age than does the average interval between successive satisfactions. This relationship is illustrated by the data on sleeping and eating presented in Tables 1 and 2. Therefore, intensity increments should decrease more sharply with development when estimated by the reciprocals of the range of intervals than when estimated by average intervals between successive satisfactions. The decreasing ratios in the third column of Table 3 show that this is the case for our sample.

We have assumed that n , the number of motives that are distinguishable within a person, increases with the person's development. This should hold for any sample of n as well. The motive to be in the company of people, for example, should give rise to an increasing number of varieties or derivatives with age. Hence the number of different people with whom a person has associated or, more operationally, the number of such people that he can list within a given period of time should increase with development or age.

This relationship was tested with another sample of subjects of the same three age groups (again high school sophomores, high school seniors, and college juniors, of both sexes, approximately matched as to socioeconomic and ethnic background) who were given five

TABLE 4
SPONTANEOUS RECALL OF FRIENDS
AND ACQUAINTANCES

Subjects (age groups)	N	Average Number of "Friends or Acquaintances"	
		Listed in 5 minutes	Present in Class
15 years	22	36.6	4.6
17 years	20	39.0	2.3
20 years	32	54.4	3.6

minutes in which to list the names of any friends and acquaintances they could think of. First names or nicknames were acceptable as long as the subjects could afterwards identify for themselves whom they had thought of. After counting the number of names in their lists, subjects were asked to state how many of those named were present in class. Table 4 shows the results, which confirm significantly the postulated trend. Deduction of those of them who were present in class and may have been accidentally looked at rather than thought of would make the trend even more pronounced.

The motive to be in the company of one's most intimate friends is a sample of the many derivatives and substitutions continua arising from the motive to be in the company of people. On the assumption that intensity increments of motives should decrease with age, even in the case of such a sample, we should expect that longer separation and a greater variability of separations from one's most intimate friends can be tolerated as a person develops. In order to test this, the same three groups of subjects were asked to list three most intimate friends and indicate the average, the longest, and the shortest separation from them for the past seven months (i.e., from September through March). The average lengths of separations were used directly, whereas the ranges of separations were computed from the longest and shortest separations indicated by each subject. For technical reasons those friends listed from whom they had been separated for all seven months were excluded from consideration. There were no such cases with the 15-year-olds, two with the 17-year-olds, and seven with the 20-year-olds. Tables 5 and 6 show the results. They confirm significantly the postulated

TABLE 5
AVERAGE DURATION OF SEPARATION FROM FRIENDS
(Means of estimated averages)

Subjects (age groups)	N	Average Duration of Separation (in days) From Most Intimate Friend Listed:			Total Mean of All Three	Range of All Three
		1st	2nd	3rd		
15 years	22	7.0	9.1	10.4	8.8	3.5
17 years	20	8.8	12.5	13.7	11.6	4.8
20 years	32	9.7	17.6	24.2	17.2	14.5

TABLE 6
RANGES OF SEPARATIONS FROM FRIENDS
(Mean differences between estimated longest and shortest intervals between get-togethers with three most intimate friends)

Subjects (age groups)	N	Range of Separation (in days) From Most Intimate Friend Listed:			Total Mean of All Three	Range of All Three
		1st	2nd	3rd		
15 years	22	13.3	13.8	22.1	16.4	8.8
17 years	20	21.6	16.0	32.4	23.3	16.4
20 years	32	18.1	32.7	43.7	31.5	25.6

TABLE 7
DURATION OF PERIODS OF CONSECUTIVE SCHOOL
WORK OUTSIDE OF CLASS
(Means of estimated average and maximum periods)

Subjects (age groups)	N	Number of Consecutive Work Hours	
		Average	Maximum
15 years	22	1.6	4.1
17 years	20	1.9	5.7
20 years	32	2.5	6.4

trend. The average separations from three most intimate friends as well as their ranges do increase with age. Even the variety among the three most intimate friends as indicated by the ranges of the means shown in Tables 5 and 6 increases with age (see last column).

Let us see whether a sample of the "free energy portion" of a person's motives, i.e., of the not-yet-recurrent Motives m , $m+1$, $m+2$, ..., $n-2$, $n-1$, n , also reflects the developmental trend postulated. Intensity increments of these motives should decrease with development, although the very fact that they have still to be satisfied for the first time makes them difficult to appraise. However, the work done for the sake of such remoter goals, or for any one

of them, should reflect the decrease of intensity increments. A person's periods of consecutive work or, put differently, the periods that he can tolerate involving delay of other satisfactions not inherent in work, should grow in length and variability with development. Hence the average as well as the largest number of consecutive hours spent at work for school during the past seven months should increase with age among our three age groups. Table 7 shows significantly that they do.

Substitution Effects

One of the various complications that hamper the automatic use of the formula being proposed is the fact that motives can do something for each other. They form substitution continua. It has been suggested that the formula can be legitimately applied to a motive only if at least the most relevant substitute motive is considered with it jointly. Following this suggestion, let us consider eating and smoking, which should be on a common continuum at least for some of their aspects. A person who smokes more may eat less, and vice versa. But even sleep may be affected. Smokers claim that cigarettes can often postpone sleep for a while. So even sleep and smoking should have something in common, be it no more than deeper breathing. In order to test these views, a group of 80 college students of an average age of 20.4 years was divided into 49 smokers (reporting average intervals and ranges of intervals between successive cigarettes of 2.1 and 14.7 hours respectively) and 31 nonsmokers. Both groups were asked to indicate their average, largest, and smallest intervals between successive occasions of sleep and meals (including snacks, but excluding the night intervals from consideration) for the past two months.

TABLE 8

SUBSTITUTION EFFECTS

(Means of estimated average intervals between satisfactions, and of range between minimum and maximum intervals)

Subjects	N	Sleeping (in hours)		Eating (in hours)	
		Avg. int.	Range	Avg. int.	Range
Smokers	49	16.3	12.6	4.5	8.0
Nonsmokers	31	16.0	11.1	3.9	6.2

Table 8 shows the results. The average intervals between successive periods of sleep and meals as well as their ranges were significantly different for smokers and nonsmokers. Smokers slept and ate somewhat less and could tolerate greater variability of either. They appeared to be slightly more "mature" when appraised by eating alone, and, apparently, smoking enabled them to appear so.

Development in an Infant

Let us see whether single motives also reflect the developmental trend of decreasing increments in an individual case. Would the postulated trend for our indices of intensity increment hold for eating and sleeping in the development of an infant? In order to test this, the sleeping and feeding schedules of a breast-fed baby weaned at the end of five months were recorded for five consecutive days in the baby's first, third, tenth, and thirteenth month of life. The means of these values are presented in Table 9.

The results confirm significantly the postulated trends. Average intervals between successive periods of sleep and successive feedings as well as their ranges do increase with development. Comparison of Table 9 with Tables 1 and 2 may make one wonder whether the infant studied was not more mature in eating than the 15-, 17-, and 20-year old. He shows a larger average interval and range of intervals at 13 months than they do. Such is not quite the case, however, since the night intervals between meals were not included in the ratings of the older subjects, whereas with the infant they were. Including the night interval would

TABLE 9

EATING AND SLEEPING IN THE DEVELOPMENT OF AN INFANT

(Mean interval between satisfactions, and mean range between minimum and maximum intervals, for five consecutive days in each month)

Subject During	Sleeping (in hours)		Eating (in hours)	
	Avg. int.	Range	Avg. int.	Range
1st month	0.8	0.4	2.8	2.4
3rd month	1.4	1.0	3.8	3.0
10th month	3.5	2.6	4.2	7.8
13th month	4.8	3.2	4.4	9.8

have raised the average intervals of all three groups by more than an hour, and their ranges by considerably more.

Additional Evidence

It was possible to secure a few additional data on the ontogenetic development of as primitive a "motive" as breathing. The author happened to record in skimpy diaries kept through several years of his youth, and for reasons partly forgotten, the lengths of time he could stop breathing at ages 11, 14, 16, and 21. The measures recorded were 70, 92, 135, and 175 seconds respectively, which reflects the developmental trend to be expected. It should be noted, however, that breathing is, by degree, quite different from all other motives mentioned. It is constantly being satisfied with anyone alive, and something like the average or maximal interval between successive inhalations depend, among other things, on such primitive constitutional givens as basal metabolism or the size of one's chest. This should be pointed out, since breathing would weigh heavily if it were included in any pool of motives. Its intensity increment for an adult is about 3000 times larger than, e.g., the intensity increment of eating. The author would not be surprised, though, if a sufficiently complex treatment of breathing should yield as worthwhile an indicator of motivational development as sleeping does. In fact, he has some inconclusive evidence that the ratio of maximal over minimal intervals between successive inhalations (in which the latter was obtained from the number of inhalations per minute after climbing rapidly three flights of stairs) does increase with the level of motivational development or ego strength as established by other measures, at least up to 20 years of age.

Finally the substitution continuum "genital satisfaction" was studied with two separate groups of ten and nine male subjects. Masturbation and heterosexual intercourse were chosen as samples. The subjects had to rate in retrospect their average, largest, and smallest intervals between successive satisfactions for different years in their development. Average intervals between successive satisfactions as well as ranges tended to increase with development, although this was true for hetero-

sexual intercourse only after marriage presumably because society places restrictions on premarital relations. Another finding may perhaps be of greater interest. Masturbation plus heterosexual intercourse treated indiscriminately as genital satisfaction and rated by the subjects for the year following their first (as it happens, premarital) heterosexual intercourse showed greater average intervals and ranges of intervals between successive satisfactions than masturbation alone during the previous year. This means that although a "new" motive, heterosexual intercourse, has been added, the developmental trend of increasing intervals and ranges of intervals between successive genital satisfactions persists. It also means that the power (i.e., the capacity to reduce motive intensity) of this new motive is greater than that of other motives assembled on the continuum. Clinical evidence confirms that of course. Masturbation tends to become rarer and more irregular as soon as heterosexual intercourse has begun. It does so even during that period of time in which the adolescent has found a love object and not yet begun to have heterosexual intercourse.

DISCUSSION

When does this end? When, on the whole, do the average intervals between successive satisfactions of motives and their ranges stop increasing and their intensity increments stop decreasing? And if there is such a thing as a decline with aging, they must. I have no definite answer. Yet somewhere between birth and death there should be a reversal of this trend. Perhaps in the third decade of life, or in the sixth. For any particular motive, however, this may happen much earlier or later. In fact, different substitution continua may develop at a different pace throughout a person's development. Apparently the developmental pace may change, provided that opportunities to satisfy the motive in question change first. A person may be motivationally underdeveloped in his contact with physical things, but well developed in his contact with people. Or he may have been drastically disappointed by people and "regressed" in this area of his world, but developed in the other. This is why a sample of a person's motivational de-

velopment should be taken from motives in different substitution continua.

Is not something like friendship really beyond the reach of all attempts at quantification? Well, certainly not in some respects, and I would not underestimate those. If I know that a wife and a husband, of their own accord, see each other only once or twice a year, or that two friends, although separated for several years at a time, have always kept up correspondence, I know something quite essential about their relationship, no matter how much I do not know. In fact, I would not want to miss that information on any accounts, if I can have it.

"All right," the reader many say, "but take then your sweeping assumption that all motives increase in intensity from the moment they have been satisfied to the moment they are being satisfied again, and that satisfaction reduces that intensity. Is that really true of complex motives? Do we not want to see more and more of a good friend or beloved, the more we have tasted of him or her?" Yes, but then we may not have had yet all we wanted. We may have "tasted," but not "eaten." Another intriguing thing about people is that they offer innumerable aspects of satisfaction which sometimes cannot be exhausted even on a thousand successive contacts. Potentially, people are the most complex of all conceivable objects. So, while things are going well, we would by definition want to see more and more of them. On the other hand, there is no person, no matter how dear, of whom we could not get tired temporarily, at least in some respects. Suppose our spouse suddenly decides never to let go of us even for a minute. She or he might become oppressive before long.

Have I then treated complex motives as if they were primary drives? If the major difference between the two kinds of forces is complexity, even the most complex motive must have something in common with primary drives. This is what a general theory of motivation must assume, if it is not merely the pretense of a theory. If I had not treated them somewhat alike, I believe I should have.

One may wonder whether I have not talked about actual events rather than motives. But how can we talk more safely about motives than in terms of actual events? If a person has

taken a meal, visited a movie theater, or even married a certain person, can we not be sure of one thing: a motive to eat, a motive to visit a movie theater, or a motive to marry that person, respectively, was one of the determinants that led to that event? This is not to deny, of course, that there are other determinants too.

Most of our data are based on self-ratings. Can they be relied on? I think so on two grounds. First, the subjects had no idea of the genetic nature of the studies, although they were asked to use pseudonyms rather than their own names for identification. This may have conveyed that the particular individual was not the focus of interest. Second, as indicated, the self-ratings concerned actual events, and very trivial ones at that. They involved no special, and possibly misleading, psychological concepts on their part.

Of the two measures used, the range of intervals between successive satisfactions of a motive should be more sensitive than the average interval. In the case of self-ratings, the range is based on the subject's estimate of his longest and shortest period of abstinence from the satisfaction of a motive over a given period of time, and both of these are likely to be relatively impressive events. The subjects should be able to rate them with greater validity than the average period of abstinence. The latter is the result of some semiarticulate computation or of an intuitive guess based on an altogether tacit computation, and that means greater possibilities of error. In accordance with this it was found that the increases with age of the ranges of intervals between successive satisfactions tended to be more significant than those of the average intervals.

The reader may have tried to account for the empirical data in other terms. I have no objection. In fact, I could do it myself. However, I think I can do it more consistently and economically with the formula suggested, and that was one of the reasons for suggesting it. The pieces of evidence should not be judged on their own merits, but rather on their compatibility with the general context in spite of their relative heterogeneity. I would not think of them as conclusive. Yet they illustrate and confirm in a variety of ways what appears to me to be conceptually conclusive.

Is the formula suggested not altogether trivial? If waking periods increase in length and variability with development, must there not be more things to fill them? Yes, but why should waking periods increase in the first place? And why should any one of the things that fill them tend to increase and vary its abstention spans in similar ways? Not that the formula explains it; it postulates the generality of this relationship. In addition, these recurrent satisfactions of distinguishable motives can fill waking periods of a given length and variability in many different ways with different people and/or under different sets of opportunities. What is more: none of these satisfactions involves one motive only. Its occurrence has multiple determination. And all of these satisfactions can do something for one another to various degrees ranging from zero substitution value to its maximum (which is satisfaction of the identical motive).

If we want to distinguish explicitly substitution continua among a person's motives, the basic formula would have to be re-written as follows:

$$\sum_{i=1}^{n_1} \epsilon_i + \sum_{i=1}^{n_2} \epsilon_i + \dots + \sum_{i=1}^{n_{p-2}} \epsilon_i + \sum_{i=1}^{n_{p-1}} \epsilon_i + \sum_{i=1}^{n_p} \epsilon_i = \sum_{i=1}^{n_p} c_i = C$$

In this formula, C is the sum of intensity increments of all motives assembled on a given substitution continuum, say, oral stimulation and manipulation; p is the number of substitution continua on which n motives (in which

$n = \sum_{i=1}^p n_i$) are assembled. In words: the sum of all sums of intensity increments of motives assembled on all of a person's substitution continua equals C . If there were as many substitution continua as motives ($p = n$, but also $n_p = n$), the motives would be unrelated to each other. If every motive appears on every substitution continuum, the number of distinguishable motives becomes the product of the number of motives that can be identified on any one continuum and the number of substitution continua ($n_p p = n$). This would reflect the most intimate interrelationship of motives possible. Obviously we will not find

this degree of interrelationship of motives or this degree of "integration" in reality, no matter how high we have a person climb on the ladder of maturity. We can expect, however, that a person with a given n will be the more mature, the greater the number of motives that are represented on more than one substitution continuum, and the greater the number of substitution continua on which they are represented. Hence, both n_p and p tend to grow with a negative increment relative to n as a person develops, so that n_p will approach $\frac{n}{p}$ and p will approach $\frac{n}{n_p}$, both, of course, without reaching either. We will learn something about a person's motivational maturity if we inspect, e.g., on how many of a given sample of substitution continua a given sample of motives appears. A person who satisfies the motive to make music by playing records, singing in a choral society, playing the violin in an orchestra, and writing compositions, is likely to have developed further in this area than a person who just plays records. Suppose, then that his motive to engage in conversations may be satisfied by talking about music and musical events, about politics, history, art, and problems involved in maintaining a family, a house, and a car, whereas the other may be talking only about musical events (and even these could be restricted to record releases) and the maintenance of house and car. And suppose that in home contacts with friends, he may be dining, drinking, chatting about current events, listening to music, and even performing some, singing operas together and playing cards, whereas the other would only be listening to music, chatting about current events (of a more restricted nature), and drinking. It is likely that the first man is more mature, not only in one area, but generally in his motivational development than the second.

C was assumed to be a quantity characteristic of the individual. But changes in C are conceptually possible. The reader should be warned, though, against thinking that changing moods, shifts from high to low feelings or vice versa, are evidence of changes of C . These moods or feelings are usually linked with specific changes of opportunity to satisfy motives, or with satisfactions themselves. There is still another consideration: the rela-

tionship between intensity increments of motives and C may be a more complicated one to begin with. Possibly it is the squares of intensity increments of a person's motives that add up to C . What speaks in favor of our initial formula in which C is taken as an intra-individually invariant quantity, however, is its simplicity, and none of our empirical tests have contradicted it. The precise nature of the relationship between C and n remains to be clarified. We have assumed no more than that n increases with development, but it is likely that the rate of increase is, in principle, a function of C and of an individual's chronological age at the time of inspection, perhaps a function of their product.

If, moreover, satisfaction of motives has any bearing on learning or cathexis (of conditions under which motives can be satisfied), such as that the amount of ongoing learning or cathexis is a function of the amount of concomitant satisfaction, C should determine the average rate of learning or cathexis characteristic of a person, or vice versa. The decline of intensity increments of given motives due to the growth of n is, perhaps, not the only manifestation of C in the course of development. Another one may be the accumulation of the results of learning or cathexis, or the growth of knowledge. The two seem interrelated even on common sense grounds, but in our terms the amount of knowledge that a person has accumulated would be a function of the level of motivation he has reached, and vice versa. As a matter of fact, the amount of knowledge may often appear a more convenient indicator of a person's general or even specific motivational development. An individual's psychological world happens to be organized not only by substitution continua, but also by objective relationships of contiguity and instrumentality. In the infant's mind, there is an order such as mother's breast, the bottle, his own hands, the pacifier, etc., but also another one like mother's breast, mother's hands (that can touch him or pass the bottle), mother's face (that can smile), mother's dress (which may change), etc. The larger the number of identifiable motives for which mother has been learned to be instrumental or, in other

words, the greater the number of aspects of her that have been cathected, or the greater the "knowledge" of mother, the farther advanced will the infant's motivational development tend to be.

On this reasoning, we should be able to infer a person's theoretical C from his rate of learning or cathexis, and the latter can be tested. We could then determine the extent to which empirical approximations to C on the basis of our formula fall short of his theoretical C . The deficit would be the result of restrictions or losses of opportunity to satisfy motives, and would provide an index of the extent to which substitution of motives by force, or repression, has hampered his motivational development. Is all this of any relevance for clinical psychology or psychiatry? I think so. I think that the clinician is implicitly proceeding by all these considerations. I have only tried to articulate some of the aspects involved in the appraisal of an individual's motivational development or "ego strength." What may be more, all aspects of the formula, if used with good sense and caution, can themselves be useful for clinical purposes. Repeated inspection of a few motives such as sleeping, eating, smoking, or even reading, but also engaging in relationships with people in general, of the opposite sex in particular, or being in the open air, may tell us whether a person is continuing to grow, beginning to slow up, or even regressing. Such data may suggest psychopathology before a person shows any more conspicuous clinical symptoms. The rates at which specific neurotic symptoms recur may also be viewed from a similar perspective.

SUMMARY

The simplest version of a comprehensive formula for the quantitative treatment of human motivation and its development was suggested. It can be handled in practice by appropriate sampling and approximations. The major aspects of the formula were subjected to empirical tests. The formula proved feasible in all of the aspects investigated.

Received September 25, 1957.

SEQUENTIAL DEPENDENCIES IN PSYCHOTHERAPY¹

FRANK AULD, JR. AND ALICE MARSDEN WHITE
Yale University

AS ONE studies psychotherapy interviews, one is struck by the lawfulness and interconnectedness of the events in them. It seems obvious that a patient, once launched on a theme, is likely to continue on that theme; it seems that various forms of resistance are equivalent, so that one kind can replace another; and it seems that interventions by the therapist occur at predictable times and have predictable effects. Is there any way of transforming such impressions into quantitative statements which would be amenable to tests of statistical significance and other techniques of quantitative measurement?

To make a quantitative description of psychotherapy, an investigator must assign each unit in the interview to one of a set of categories. In the research reported here, the sentence was chosen as the unit to be classified. Each sentence, therefore, had to be assigned to one of the categories of our content-analysis system. The system of classification used and the rules for categorization were developed by Dollard and Auld (6). The categories referred to in this paper are only a few of the more than 60 in the system.

Having assigned each unit to one of the categories, one can then study various questions. Four questions are considered in this paper.

1. Does the patient's speech hang together, i.e., is the patient likely to continue with another unit belonging to the same category?

Because the patient's behavior hangs together and because he is motivated by learned drives to be intelligible and logical, we expect his talk or silent thought about a topic to persist. For instance, if the patient in one unit is talking about his hostile feeling toward his wife, he may be expected to continue to talk about this in the next unit. Thus, a unit scored as "hostility" should often be followed by another unit scored the same way. A

similar prediction would be made about any content category.

2. Are various forms of resistance equivalent as shown by their occurrence in units that follow one another?

If the patient resorts to silence in a case conducted under the free-association condition, he is breaking the free-association rule. Such behavior is believed by psychoanalysts (5, pp. 103-106) to be resistant. If silence is resistant, one might reasonably expect that communications of the patient preceding or following silence would also be resistant; that is to say, if the patient is resistant at a particular time, it will be shown at one moment by silence and at another moment by resistant talk. According to psychoanalytic theory, therefore, silences should often be preceded and followed by resistant talk.

Even when the free-association rule has not been laid down, it is reasonable to consider silences as resistant. According to psychoanalytic theory, communication from patient to therapist is necessary to progress. When the patient is silent, he is not communicating verbally with the therapist; therefore, he is failing to act in a way that could advance the work.

Whether silence is in fact correlated with verbal resistance should, therefore, test the correctness of analytic views about silence. It should be said, however, that psychoanalysts would consider the context in which the silence occurred when deciding whether it was resistant. Our study of silences does not include this refinement; we are only testing whether *in general*—neglecting nuances of context—silences function as resistance.

3. Does the patient's activity affect the likelihood that the therapist will intervene and, if he does intervene, the kind of intervention he will make?

According to psychoanalytic theory, the therapist ought to intervene when the patient is resistant, and the therapist's intervention should be an interpretation of the resistance. Resistances must be dealt with to make it possible for the patient to continue to learn

¹ This investigation was supported by a research grant, M-648 (C3), from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

more about himself, in other words, to continue making the unconscious conscious (8, pp. 61-69).

Thus, a therapist acting on psychoanalytic principles would be expected to intervene more often after resistant talk than after non-resistant talk, and he would be expected to intervene by interpretation. The more skilled the therapist, the better able he should be to identify resistances quickly and to interpret them promptly.

4. Does the kind of intervention made by the therapist affect the subsequent activity of the patient?

Any adequate theory of psychotherapy must account for the effect of the therapist's interventions on the patient. Psychoanalytic theory, attempting to do this, describes some interventions as "interpretations," that is, as attempts by the therapist to label unconscious responses of the patient (5, p. 82; 7). Other interventions are called "support," "guidance," "reassurance," "questions." Analytic theory says that an interpretation of a resistance, if apt, interrupts the resistance and allows the patient to resume his task of learning about himself. Client-centered therapists, on the other hand, believe that interpretations are likely to evoke resistance and interfere with the therapeutic process. According to these therapists, clarification of feeling rather than interpretation is the preferred technique (10).

In this paper, results throwing light on these four questions are presented. Our aim, however, is somewhat more general than to answer these questions; it is to demonstrate the value of studying *sequential dependencies* in psychotherapy, or, in other words, the value of studying what-follows-what.

The method, in general, is to study whether the kind of score assigned to Unit n affects the likelihoods of various scores being assigned to Unit $n + 1$. For example, if one knows that the patient has spoken a sentence classified as "resistance," does that help to predict what kind of unit will come next?

We make no claim to originality in using content analysis to study psychotherapy (2) or even in studying what-follows-what. However, we do believe that no one has previously worked with psychotherapy data in the form of transition probabilities.

METHOD

Material. Four psychotherapy cases, carried by four different therapists, were studied. Two of the therapists (who treated Patients A and B) had been fully trained in psychoanalysis and had considerable experience; the other two, having taken an introductory course in psychotherapy, were having their first experiences as psychotherapists. All of the patients had applied for psychotherapy to a psychiatric outpatient clinic, and their treatment was carried on under the auspices of the clinic.

Recordings were made of all interviews with the knowledge and consent of the patients. Patient A had 73 sessions, Patient B had 17, Patient C had 12, and Patient D had 34. Although every interview was carefully transcribed, not all of the hours were scored. Eighteen hours from Case A, 9 from B, 11 from C, and 12 from D were analyzed. The original purpose in scoring these interviews was to try out the developing content-analysis system. Therefore, the interviews chosen for scoring are not ideally representative of Cases A and D but come instead from the earliest parts of these cases. The hours selected from Case B are representative, since all odd-numbered interviews were scored. All hours but one of Case C were scored, the first being omitted because the scorers did not wish to deal with the special features of an initial interview.

All the cases may be described as examples of psychoanalytically oriented psychotherapy. In Cases A, C, and D, the therapist insisted on the free-association rule; in Case B, the therapist did not, making the interview more like an ordinary conversation. The therapist of Case B also talked much more than the other therapists, so that his utterances occupied about a third of the total time. Therapist A's utterances took up about a fifth of the time, and Therapists C and D talked less than A.

None of these cases was an outstanding success, and one—Case C—must be considered a definite therapeutic failure. Although not outstandingly successful, Cases A, B, and D were, in our opinion, conducted with at least average skill, compared to the general run of cases in the psychiatric outpatient clinic.

TABLE 1
RELIABILITY OF CONTENT ANALYSIS

Category	Tau*
For patient's sentences:	
Anxiety	.49
Dependence	.37
Hostility	.57
Love	.39
Mild Agreement	.83
Resistance	.58
Sex	.75
Social Mobility	.75
For therapist's utterances:	
Demand	.65
Interpretation	.59
Reward	.73

* All coefficients shown are statistically significant at the .001 level.

Content analysis. Each transcribed interview was divided into sentences according to the rules developed by Auld and White (3). Then two psychologists² independently scored each sentence of the interviews included in this study.

In evaluating the reliability of this scoring, one must consider each unit separately, because the scores are to be used in studying the relationships between *single units* and the next-following units.

Thus, product-moment correlations between the total counts for various categories for whole hours are irrelevant.³ To measure agreement on separate units, a rank-order correlation, tau (9, 1), was computed for each of the categories used in this paper. The coefficients are shown in Table 1.

Counting of sequences and computation of probabilities. The various sequences that were to be studied were counted, then were expressed as conditional probabilities. For example, the $H \rightarrow H$ sequence (score of H on Unit n , followed by score of H on Unit $n + 1$) was counted. The number of H scores was also counted. Then the conditional probability of an H on Unit $n + 1$, given that Unit n was scored H , was computed. This probability is, of course, the number of $H \rightarrow H$ sequences, divided by the number of H scores.

The possible scores for each unit can be considered "states" of the psychotherapy situation. These states are mutually exclusive and exhaustive, because the content analysis assigns each sentence to one and only one category. The conditional probabilities can be called *transition probabilities*, i.e., probabilities of the passage of the psychotherapy situation from State A to State B, if one wishes to adopt the terminology of mathematical statistics (4).

RESULTS

Coherence of patient's speech. To study the likelihood of persistence of categories, we computed the transition probabilities of the categories for each of our four cases. It was discovered that the categories do, indeed, persist. For instance, the likelihood of a sentence scored Sex is greater after Sex sentences than after other sentences, and the likelihood of a sentence scored Hostility is greater after hos-

tile sentences than after nonhostile sentences. Such a result was obtained for all of the categories studied: Anxiety, Dependence, Hostility, Love, Resistance, Sex, Silence, and Social Mobility. All of these differences are statistically significant and quite large. Furthermore, these differences were found in every case studied.

A typical finding is that for the Hostility category. The chances are 71 out of 100 that Sentence $n + 1$ was scored Hostility if Sentence n was scored Hostility. However, if Sentence n was not scored Hostility, the chances are only 3 in 100 that Sentence $n + 1$ was scored Hostility.

It occurred to us that this finding might be attributable, in part, to the tendency of a scorer to score a whole sequence of sentences the same way. In other words, the result would demonstrate a coherence in the scorer's behavior, not in the patient's. To exclude any effect of the scorer's tendency to score successive sentences the same way, we therefore analyzed the data again, using the scores of different scorers for the two sentences of each sequence. If Scorer A's judgment was used in classifying Sentence n , Scorer B's was used in classifying Sentence $n + 1$. As expected, this procedure cut down the size of the differences found; but the differences remained, were still very large, and were still statistically significant. For example, it was found (when the material was analyzed in this way) that if Sentence n was scored Hostility, the probability was .51 that Sentence $n + 1$ was Hostility; but if n was not Hostility, the probability was only .07 that $n + 1$ was Hostility.

Equivalence of silence and resistant talk. Silence did, indeed, occur more often after resistant talk than after nonresistant talk (see Fig. 1). The probability was .38 that a silence of at least 5 seconds would occur after a sentence scored as resistant; it was only .04 that a silence would occur after a nonresistant sentence. This difference is statistically highly significant. Furthermore, silences were likely to be followed by resistant talk. Dividing each silence into 5-second units,⁴ one finds that the silence would continue for another 5 seconds

² John Dollard and Frank Auld, Jr.

³ Such hour-total correlations, we have found, tend to be higher than correlations computed from scores on single units; for instance, in Table 1 the median tau coefficient (based on single-sentence comparison) for patient's categories is .58; the median r (based on hour-total comparison) for the same categories is .87. The larger size of r 's can be attributed to compensating errors in the hour-counts of the two scorers. For instance, Scorer A calls Units 1 to 10 "resistance," while Scorer B calls them "anxiety"; but Scorer B calls Units 53 to 62 "resistance," while Scorer A calls these "hostility." Considering both sequences of units together, the scorers agree that 10 of the units are "resistance," even though they do not agree on the scoring of any of the separate units.

⁴ Five seconds was chosen because the patient could utter one sentence in this period of time.

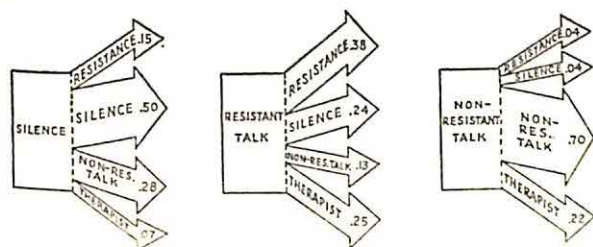
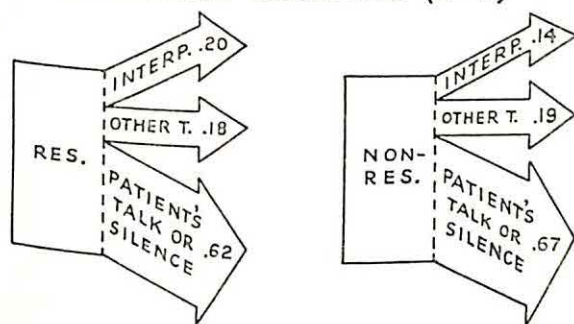


FIG. 1. EVENTS FOLLOWING SILENCE, FOLLOWING RESISTANT TALK, AND FOLLOWING NONRESISTANT TALK

EXPERIENCED THERAPISTS (N=2)



APPRENTICE THERAPISTS (N=2)

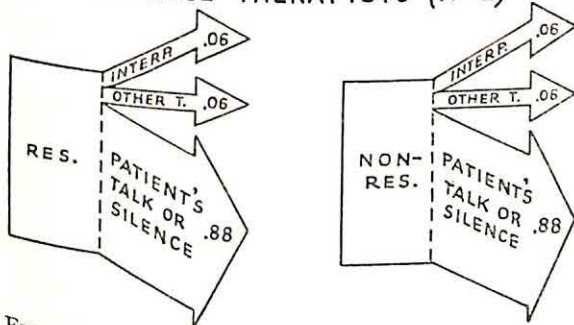


FIG. 2. THERAPIST'S RESPONSE TO RESISTANT AND TO NONRESISTANT TALK OF PATIENT

50% of the time. A resistant sentence occurred in 15 out of 100 cases, and nonresistant talk or an utterance by the therapist occurred in 35 out of 100 cases. The .15 probability of resistant talk after silence may be compared to the .04 probability of resistant talk after nonresistant talk. The difference, of course, is highly reliable.

These findings are what would be expected if silences were equivalent to resistant speech.

Effect of patient's activity on therapist's behavior. The two apprentice therapists were not more likely to intervene after resistant talk than after nonresistant talk. If they did intervene, they were equally likely to make an

TABLE 2
LIKELIHOOD OF VARIOUS RESPONSES BY PATIENT AFTER INTERVENTIONS BY THERAPIST

Patient's Response	Therapist's Intervention	
	Interpretive	Noninterpretive
Resistant talk	.21	.20
Nonresistant talk (excluding Mild Agreement)	.24	.54
Mild Agreement	.49	.21
Silence	.06	.05

interpretation or to make a noninterpretative remark. The interventions of these beginners, it can be seen, do not fit very well the analytic description of correct technique.

The experienced therapists, on the other hand, were more likely to interpret resistance. Figure 2 presents the over-all results. Therapist A was more likely to intervene after resistant talk than after nonresistant talk (the probability figures are .24 and .18, respectively; the difference is statistically significant). Therapist B was about equally likely to intervene after resistant or nonresistant talk (.50, .47). But if he did intervene after resistance, he made an interpretation more often; 38% of his interventions after resistance and 29% of them after nonresistant talk were interpretative. This difference is statistically significant.

It is apparent that the experienced therapists were better able than the inexperienced ones to identify quickly and to deal directly with resistance. Since interpretation is believed to be the method whereby the therapist helps make the patient's unconscious conscious, by interpreting resistance these therapists were acting in accord with theory in the interventions that they made.

It should also be noted that the experienced therapists talked more than the apprentice therapists in the cases studied. However, this finding should not be taken as evidence that skillful therapists must talk as much as Therapists A and B did in the cases studied here. In other cases that we have studied, a skilled therapist acted very adaptively without having to talk much.

Effect of therapist's intervention on patient's behavior. We made a preliminary study of the effect of interpretations by noting what the

next sentence of the patient was after an interpretation and after noninterpretative interventions. The results are given in Table 2. It can be seen that, if one may judge by what the patient says immediately afterward, interpretation did not produce a great increase in resistance. The probability of the patient's saying "Umhum" or "Yeah" was, however, much greater after an interpretation than after other interventions. (The difference is highly reliable.) No doubt this is so because most interpretations call for some answer from the patient, some indication that he agrees or disagrees. It should be emphasized, however, that we do not suppose that the patient's Yes indicates more than polite agreement—"I heard you." Such a reply cannot at all be considered evidence that the patient has accepted and adopted the interpretation.

Further studies on the effect of various kinds of intervention by the therapist are planned; it is hoped that they will throw some light on the controversies concerning what moves psychotherapy ahead.

SUMMARY

As a demonstration of the value of studying sequential dependencies in psychotherapy, data are presented bearing on four hypotheses: that the patient's speech is coherent, that silence is equivalent to resistant speech, that analytic therapists are more likely to inter-

vene after resistances, and that interpretations by the therapist produce an upswelling of resistance. The first three hypotheses were borne out, but the last one (derived from client-centered theory) was not. The authors believe that the results obtained justify wider use of this method of studying psychotherapy.

REFERENCES

1. AULD, F., ERON, L. D., & LAFFAL, J. Application of Guttman's scaling method to the T.A.T. *Educ. psychol. Measmt*, 1955, **15**, 422-435.
2. AULD, F., & MURRAY, E. J. Content-analysis studies of psychotherapy. *Psychol. Bull.*, 1955, **52**, 377-395.
3. AULD, F., & WHITE, ALICE M. Rules for dividing interviews into sentences. *J. Psychol.*, 1956, **42**, 273-281.
4. BUSH, R. R., & MOSTELLER, F. *Stochastic models for learning*. New York: Wiley, 1955.
5. COLBY, K. M. *A primer for psychotherapists*. New York: Ronald, 1951.
6. DOLLARD, J., & AULD, F. Development of quantitative methods for detailed study of psychotherapy. Progress report to National Institute of Mental Health, Public Health Service, 1956. New Haven: Authors, 1956.
7. DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.
8. FREUD, S. *An outline of psychoanalysis*. New York: Norton, 1949.
9. KENDALL, M. G. *Rank correlation methods*. London: Griffin, 1948.
10. ROGERS, C. R. The development of insight in a counseling relationship. *J. consult. Psychol.*, 1944, **8**, 331-341.

Received October 1, 1957.

A COMPARISON OF MENTAL RETARDATES AND NORMALS ON VISUAL FIGURAL AFTEREFFECTS AND REVERSIBLE FIGURES¹

HERMAN H. SPITZ AND LEONARD S. BLACKMAN

Edward R. Johnstone Training and Research Center, Bordentown, New Jersey

THE question of whether mental retardates exhibit more rigid behavior than normals is far from settled. Lewin (12) reported that retarded children display more persistence in a single activity than do normals of approximately equal CA, although he found no significant difference between normal and retarded children in total satiation time. A more recent study (19) has indicated that retardates, although much more influenced by the support or nonsupport of the examiner, persist in monotonous tasks for longer periods of time than do normals of equal MA. In an experiment measuring the performance of equal MA groups on discrimination learning tasks, Stevenson and Zigler (15) reported that retarded children with a mean CA of 10.9 years, as well as retarded adults, performed as well as normal children with a mean CA of 5.0 years. Such results argue that low MA retardates are no more rigid than normals of equal MA. However, because the normals of that study were chronologically only 5 years of age—an age at which inflexible types of behavior have frequently been demonstrated (13, 16)—Stevenson and Zigler's results do not dispel the notion that, in general, retardates may be described as rather rigid individuals.

The present study proposes to compare the performances of retardates and normals of equal chronological age on (a) a test presumed to measure perceptual rigidity (4) and (b) a test presumed to measure neural modifiability (18). The results of these comparisons may add to our understanding of the possible relationship between the construct of perceptual rigidity and the construct of neural modifiability.

Although Lewin used the term "satiation" in its more usual sense of behavioral boredom or fatigue, Köhler and Wallach (10) use it to describe a neurophysiological process which,

¹ This study is the first of a series designed to describe mental retardation in terms of those specific components of perception, learning, cognition, and personality which constitute the dynamic response complex referred to as intelligent or adaptive behavior.

among other things, is said to be responsible for visual figural aftereffects. These aftereffects can be demonstrated quite readily with various kinds of figures. The subject (S) is asked to fixate on a figure (Inspection Figure) for varying lengths of time, at the conclusion of which a second figure (Test Figure) is presented. The Test Figure appears to be displaced away from the area previously occupied by the Inspection Figure, and it may appear to change in size, luminosity, and depth. Displacement effects can also be found in the kinesthetic (5, 8) and auditory (11) sense modalities.

From their findings, Köhler and Wallach concluded that the presence of a figural current tends to block the further presence of figural currents in the same area. It is this satiation process that is said to account for figural aftereffects, causing the Test Figure to move *away* from the area of impedance or satiation.

The process of neural satiation is also said to account for the reversal of figure and ground, as illustrated in reversals reported during the fixation of such stimuli as the Necker cube or the Rubin Vase-Profile figure. However, Harrower's (4) modification of the Rubin figure, in which no steady fixation is required and in which consecutive changes are made on the structure of the stimuli, presumably makes her test more a measure of general perceptual rigidity than specific neural modifiability. Nevertheless, if different stages of the neural satiation process occur in all perception, then this process may play some role in the capacity to perceive both figures in a reversible figure, even though there is no fixation required of the S.

If the assumption is made that neural satiation is a physiological process which is not only responsible for certain perceptual processes, but is made manifest by these very processes, then tests of satiation are perceptual pathways to neurological capacity (14). It would be important to know whether there is

a basic neurological difference, as measured by certain perceptual tasks, in a group of mentally retarded Ss as compared with a group with average intellectual development.

The first hypothesis of the present study is that there is a difference between the experimental group of institutionalized high-grade mentally retarded boys, both endogenous and exogenous, and a control group of normal high school students of approximately the same chronological age in capacity to perceive visual figural aftereffects.

The second hypothesis is that there is a difference in the perceptual rigidity of experimental and control groups as defined by the capacity to shift figure and ground on a modified Rubin Vase-Profile test.

The third hypothesis holds that perceptual rigidity varies inversely with the capacity to satiate. That is, low perceptual rigidity accompanies high capacity to satiate, as would be expected on the basis of satiation theory as well as from the results of studies reported by Klein (6).

METHOD

Subjects. The experimental group for the Reversible Figures test consisted of 87 high-grade mentally retarded adolescent boys, ranging in age from 13 to 21 years, all residents of a short-term state training school. Of these, 37 completed the Figural Aftereffect test. The difference in number of Ss was due primarily to certain restrictions, as indicated below, in the control run of the aftereffect test. All but one of the Ss of the experimental group had been administered individual intelligence tests within a year of this study.

The control group for the Reversible Figures test consisted of 57 male students from a local high school who were closely equated in chronological age with the experimental group. Of these, 41 completed the Figural Aftereffect tests. All of these Ss had been administered the California Mental Maturity Test at some time during their high school stay.² A comparison of the mean chronological ages and IQs of the experimental and control groups is given in Table 1. Ss were chosen at random within a specified age stratum. As can be seen from Table 1, the two groups were essentially the same in chronological age, but differences in intelligence were significant well beyond the .001 level.

Tests. Two tests were administered to all subjects: (a) a modification of the Rubin Vase-Profile Reversible

TABLE 1
IQ AND CHRONOLOGICAL AGE DISTRIBUTION OF
MENTALLY RETARDED (EXPERIMENTAL) AND
NORMAL (CONTROL) GROUPS

Test	N	Mean Age (Yrs.)	SD	Mean IQ	SD
Figural Aftereffects Test					
Experimental	37	17.34	1.69	66.11*	12.08
Control	41	16.94	1.32	96.80*	10.82
Reversible Figures Test					
Experimental	87	17.08	1.80	61.52*	12.84
Control	57	16.81	1.28	95.30*	11.70

* Differences significant at the .001 level.

Figure test and (b) a Visual Figural Aftereffect test. A Rubin Vase-Profile Figure, 1½ inches high, sketched in black India ink on a plain, white three-by-five card, was presented to each S, who was then asked, "What does this look like?" If the S gave a response to either the center (vase) or the side (profile) areas, he was further asked, "Does it look like anything else?" until he either gave a response to the opposite area or said he could see nothing more. If he voluntarily shifted to the opposite area (center to side or side to center) he was given a score of four points, and this phase of his testing was complete.

If S could not voluntarily shift, a series of modified Rubin Vase-Profile Figures was substituted for the original more ambiguous figure. These modifications, following Harrower (4), were used in an attempt to force a perceptual shift from figure to ground. If S had initially responded only to the side or only to the center area, he was presented with additional figures which were progressively more structured in the opposite area. Each of these additions presumably gave the ground area more substance as a figure, and therefore made it more likely to be perceived.

As soon as S gave a figural response to the area that had been the ground for his initial percept, the test was halted. Four points were scored for a "voluntary" shift and three, two, and one point, respectively, for shifts on the progressively more structured cards. If S could not shift even after having been presented with the final, most structured card, his score was zero. A zero score was also given to any S who could not perceive any figure on the initially presented reversible figure, a score which turned out to be applicable to only one control and one experimental S. Thus, each S could score from zero to four, depending on his capacity to shift. The lower the score, the greater S's perceptual rigidity was estimated to be.

The Inspection (I) Figures and Test (T) Figures used in the present experiment for figural aftereffect followed those reported by George (3) to be the most likely to elicit the aftereffect of change in size. All figures were drawn in black India ink on white showcase board. The I card consisted of two outline circles of unequal size, the center of the smaller circle being three inches above, and the center of the larger circle being three inches below, a small fixation dot. The diameter

² The authors wish to thank Frank Lucia for administering some of the California Mental Maturity Tests. Appreciation is also extended to Gilbert Papp, former principal of Bordentown High School, and George Dare, Superintendent of Schools in Bordentown, N. J., for their fine cooperation.

of the upper circle was $1\frac{1}{16}$ inches, and of the lower circle $3\frac{1}{2}$ inches. The fixation dot was $\frac{1}{16}$ of an inch in diameter. The white ground visible to the observer was 12 inches high and 18 inches wide. Ss were seated 72 inches from the apparatus, which was at about eye level. An adjustable chin rest was used throughout.

The T card consisted of two equal outline squares. Each was four square inches, with the center of the upper square three inches above the fixation dot and the center of the lower square three inches below the fixation dot. The T card could be dropped in front of the I card in such a way that the lower square fell in the area previously enclosed by the contours of the lower (larger) circle, while the upper square bounded the contours of the previously exposed upper (smaller) circle. According to satiation theory, fixation of the I card should result in an aftereffect in which the upper square of the T card appears larger than the lower one.

The Ss were randomly divided into Subgroups I and II for the purpose of varying the sequence of fixation periods. The experiment began for each S with a control test. The T card was exposed, S was told to keep his eyes on the fixation dot and asked, "With your eyes still on the dot, tell me right away: Is one box bigger than the other or are the boxes the same size?" These instructions were given before every judgment except that the question, "Are the boxes the same size?" alternated position with the question, "Is one box bigger than the other?" in a prearranged manner. If S reported during the control test that the squares were unequal, he was immediately disqualified from the Figural Aftereffect experiment. This was to eliminate possible contamination of the dependent variable by poor discrimination.

If he passed the control test by reporting the two squares as equal, the I card was exposed for a prescribed length of time, with S instructed to fixate on the center dot. Each S was assured during this fixation period that he could blink his eyes as much as he wanted to, but he was reminded to keep his eyes on the dot. Ss of Group I fixated for 1 minute, after which the T card was immediately dropped in front of the I card and a size judgment requested. As soon as a judgment was made, the I card was again exposed, but this time for a $1\frac{1}{2}$ -minute fixation period, after which another judgment of the size of the squares on the T card was requested. The apparatus was then screened from view and a 2-minute rest period ensued. After 2 minutes, S resumed his previous position, and another judgment of the T card was made. This last judgment was a test of the persistence of the aftereffect.

Group II went through exactly the same procedure, except that their inspection times were 2 and $2\frac{1}{2}$ minutes respectively, and their rest period was 4 minutes.

After no less than three weeks, all Ss were retested. At this second testing, Group I went through Group II's original procedure, and vice versa. If any S failed the control test on this second testing, he was excluded from the study. Each S who completed the experiment had to go through four fixation periods of varying lengths of time and make eight judgments: two control judgments, four postfixation judgments, and two persistence judgments.

Scores were obtained by giving .25 for each post-fixation judgment at which an S reported the top square as bigger than the bottom one. The highest possible satiation score, then, was 1.00. On the two persistence trials a score of one was given for each positive satiation judgment. A total of two would be the highest possible persistence score.

RESULTS

The first hypothesis of the present study was that there is a difference between mental retardates and normals in the capacity to perceive visual figural aftereffects. This hypothesis was supported by the results (see Table 2). The mean satiation score for the control group was .591 and for the experimental group, .189. The variance of the control group, however, was significantly greater than the variance of the experimental group, producing a significant *F* value of 2.502. Since the standard *t* test assumes homogeneity of variance, an approximation *t* was computed to test the difference between the means when the variances of the two groups are significantly different (1, p. 167). The computed *t* value of 5.289 was significant at the .001 level.

To determine whether this difference might be related to etiology within the mentally retarded classification, the experimental group was further subdivided into two subgroups: an exogenous group of 16 Ss, based on psychological and neurological diagnoses, and an endogenous group of 21 Ss who showed no evidence of brain damage. Differences between the mean satiation scores of these two subgroups yielded a *t* of 1.329, which did not reach the .05 level of significance.

Tables 3 and 4 show the number of Ss of each group who satiated, and the chi square values representing the degree of relationship between intelligence and the ability to satiate at each postfixation period and after each

TABLE 2
DIFFERENCES BETWEEN THE MEAN SCORES OF
MENTAL RETARDATES AND NORMALS ON TWO
PERCEPTUAL TESTS

	Figural Aftereffect Test				Reversible Figures Test			
	N	Mean	SD	<i>t</i>	N	Mean	SD	<i>t</i>
Normals	41	.591	.405	5.29*	57	3.105	.913	
Retardates	37	.189	.256		87	2.563	.915	3.53*

* Differences significant at the .001 level, two tailed test.

TABLE 3
NUMBER AND PERCENTAGE OF "SATIATION
RESPONSES" AFTER EACH FIXATION PERIOD

	Fixation Time			
	1 min.	1½ min.	2 min.	2½ min.
Normals	26 (63%)	25 (61%)	26 (63%)	22 (54%)
Retardates	6 (16%)	10 (27%)	6 (16%)	9 (24%)
χ^2	17.91**	9.06*	17.91**	6.99*

* Chi squares significant at the .01 level, two tailed test.

** Chi squares significant at the .001 level, two tailed test.

TABLE 4
NUMBER AND PERCENTAGE OF "SATIATION
RESPONSES" AFTER TWO REST PERIODS

	Rest Period	
	2 min.	4 min.
Normals (41)	7 (17%)	6 (15%)
Retardates (37)	7 (19%)	7 (19%)
χ^2	.04 (NS)*	.26 (NS)*

* Chi square not significant at .05 level.

two- and four-minute rest period. It is obvious from these results that the mental retardates are far less able to satiate immediately after any of the four fixation periods than are the normals. Tests of the persistence of the aftereffects indicate that in those retardates who did satiate, the effects of satiation did not dissipate as quickly as they did in normals.

Another important finding is the general tendency for the mentally retarded group to require longer fixation times before satiating. In other words, the mentally retarded Ss of the present study show a limited capacity to satiate, but those who do satiate require longer fixation periods to do so, and their aftereffects tend to persist longer, relative to original satiation levels, than do the aftereffects of normals. After the second period of testing, that is, after the second postfixation judgment on a particular day, there appears to be a satiation build-up in the mental retardates. This occurs whether or not the second period falls on the day of the shorter, 1- to 1½-minute test periods, or on the day of the longer, 2- to 2½-minute test periods.

It should be noted, however, that three normal Ss, who had satiated on their last persistence test, continued to report the upper square as larger on their second control test given three weeks later, and therefore had to

be dropped from the figural aftereffect phase of the study. These were the only Ss who gave this type of result, but the possibility must remain that for them the effects of satiation persisted over a three-week period of time. That aftereffects may possibly have persisted over an extended period for only three of the Ss would suggest that this phenomenon, if it exists at all, is a rare one (9).

The second hypothesis, predicting a difference in perceptual rigidity between normals and retardates, was also confirmed. On the modified Rubin test, the mean score for the control group was 3.105, and for the experimental group it was 2.563, a difference significant at the .001 level (see Table 2). If it can be assumed that this Reversible Figure test is a measure of perceptual rigidity, then certainly the mentally retarded Ss are far more rigid than the control group. Differences between the 16 exogenous and 21 endogenous retardates on the Rubin test did not reach the .05 level of significance, suggesting that these two etiological groups perform similarly on the particular perceptual tasks under study.

Finally, a correlation of .33 between the scores on the Visual Figural Aftereffects test and the scores on the Reversible Figures test, combining the scores of both the experimental and control Ss, is significant beyond the .01 level. This correlation must be considered cautiously, since it included a heterogeneous population. In fact, using only the 41 normal Ss, a correlation of .28 falls between the .1 and .05 significance levels. Nevertheless, it lends some support to the third hypothesis that perceptual rigidity varies inversely with the capacity to satiate. The explanation for the positive correlation supporting an hypothesis of inverse relations is that a high score on the Reversible Figures test is indicative of low perceptual rigidity, whereas a high score on the Aftereffects test indicates high satiability. This correlation further suggests that there may well be a common factor underlying the capacity to shift figure and ground, even without fixation, and the capacity to perceive visual figural aftereffects.

DISCUSSION

Although the low scores of the retarded group on the aftereffect test are assumed to be due to poor satiability, there are three other

possibilities which must be taken into account.

(a) The retardates were unable to fixate as well as normals and therefore did not perceive the aftereffects. This possibility was greatly decreased, if not entirely eliminated, by having an experimenter continually observe the S in order to remind him, if necessary, to continue to fixate. If fixation was impossible, the S was dropped from the study. (b) It is possible that the retarded Ss *perceived* the lower square as smaller after fixation, but *reported* the two squares as equal. In order to assess this possibility, a study is planned in which the Ss will be conditioned to the aftereffect. In this way, the response will not be dependent upon a verbal report. (c) It may be that the mentally retarded satiated to the same degree as did the normals, but were unable to make fine size discriminations. In this regard, one additional set of data which seemed highly significant resulted from the study. Excluding those 3 Ss who may have persisted in satiating over a three-week period of time, 11 out of 52 normals inaccurately perceived two equal squares as being unequal on one of the two control tests. Of the mentally retarded, on the other hand, 44 out of 81 failed on one of the two control runs. This suggests that the mentally retarded are far poorer than normals in the ability to make size discriminations. This lead was followed a step further. The mean IQ of the 37 retardates who had successfully completed the two control tests was 66.11. The mean IQ of the 44 retardates who failed at least one of the control tests was 57.07, a difference significant at the .01 level. The mean chronological ages of these two groups were not significantly different. These preliminary data would seem to indicate that intelligence is intimately related to the capacity to make accurate size discriminations.

The results of the reversible figure experiment, indicating that retardates as a group show greater perceptual rigidity, is in accord with what might be expected from the results of the figural aftereffect experiment. The measure of perceptual rigidity used in the present study may be just one measure of a higher order trait, much as Klein's "leveling" and "sharpening" groups were isolated on the basis of their performances on particular perceptual tasks (6). This trait may, in turn, be

the manifestation of a general cortical capacity for change which Wertheimer (18) has labeled "brain modifiability" and which Klein and Krech (7) have called "cortical conductivity." If this is the case, the high-grade mental retardates of the present study exhibit a noticeable limitation in this area, a limitation that is not confined to exogenous or endogenous types (2, p. 41). These results lend support to the contention of Lewin (12) that retardates exhibit more rigid behavior than do normals, at least when matched for chronological age.

The results of the present study may be compared with those reported by Wertheimer and his associates (17, 18), who found that schizophrenics show poorer satiability on tests of both visual and kinesthetic aftereffects. The interpretation which suggests itself is that high-grade mental retardates show a resistance to cortical change similar to that found in schizophrenics. Anyone who has attempted to modify schizophrenic thinking and behavior by means of psychotherapy, and who has also had experience in teaching the mentally retarded, should have little difficulty in seeing some merit in this interpretation. Wertheimer attributed his results to a lowered efficiency of tissue metabolism in the schizophrenics.

SUMMARY

A group of institutionalized high-grade mentally retarded adolescent boys was compared with a group of normal boys, equated for chronological age, on two perceptual tasks. The results were interpreted in terms of Köhler and Wallach's Theory of Satiation. The group of retardates showed a significantly poorer capacity to satiate, as measured by a Visual Figural Aftereffect test. Tests of the persistence of the aftereffects indicated that the effects of satiation do not dissipate as rapidly in retardates as they do in normals.

The mentally retarded Ss manifested significantly greater perceptual rigidity, as defined by a modified Rubin Vase-Profile Reversible Figures test, than the normal Ss. No significant differences were found between endogenous and exogenous retardates on either test.

A significant correlation was found between scores on the Visual Figural Aftereffects test and the Rubin Vase-Profile Reversible Fig-

ures test, suggesting a common factor underlying both perceptual rigidity and limited capacity to satiate.

REFERENCES

1. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
2. GALLAGHER, J. J. A comparison of brain-injured and non-brain-injured mentally retarded children on several psychological variables. *Monogr. Res. Child Develpm.*, 1957, **22**, No. 2 (Serial No. 65).
3. GEORGE, F. H. On the figural after-effect. *Quart. J. exp. Psychol.*, 1953, **5**, 128-135.
4. HARROWER, MOLLY R. Changes in figure and ground perception with cortical lesions. *Brit. J. Psychol.*, 1939, **30**, 47-51.
5. JAFFE, R. Kinesthetic after-effects following cerebral lesions. *Amer. J. Psychol.*, 1954, **67**, 666-676.
6. KLEIN, G. S. The personal world through perception. In R. R. Blake & G. V. Ramsey (Eds.) *Perception: An approach to personality*. New York: Ronald, 1951. Pp. 328-355.
7. KLEIN, G. S., & KRECH, D. Cortical conductivity in the brain injured. *J. Pers.*, 1952, **21**, 118-148.
8. KÖHLER, W., & DINNERSTEIN, DOROTHY. Figural after-effects in kinesthesia. In *Miscellanea psychologica Albert Michotte*. Paris: Librairie Philosophique, 1947. Pp. 196-220.
9. KÖHLER, W., & FISHBACK, JULIA. The destruction of the Müller-Lyer illusion in repeated trials: II. Satiation patterns and memory traces. *J. exp. Psychol.*, 1950, **40**, 398-410.
10. KÖHLER, W., & WALLACH, H. Figural after-effects: An investigation of visual processes. *Proc. Amer. phil. Soc.*, 1944, **88**, 269-357.
11. KRAUSKOPF, J. Figural after-effects in auditory space. *Amer. J. Psychol.*, 1954, **67**, 278-287.
12. LEWIN, K. *A dynamic theory of personality*. New York: McGraw-Hill, 1935.
13. REICHARD, SUZANNE, SCHNEIDER, MARION, & RAPAPORT, D. The development of concept formation in children. *Amer. J. Orthopsychiat.*, 1944, **14**, 156-162.
14. SPITZ, H. H. The present status of the Köhler-Wallach theory of satiation. *Psychol. Bull.*, 1958, **55**, 1-28.
15. STEVENSON, H. W., & ZIGLER, E. F. Discrimination learning and rigidity in normal and feeble-minded individuals. *J. Pers.*, 1957, **25**, 699-711.
16. THOMPSON, JANE. The ability of children of different grade levels to verbalize on sorting tests. *J. Psychol.*, 1941, **11**, 119-126.
17. WERTHEIMER, M. The differential stability of schizophrenic and normal subjects: A test of a deduction from the theory of figural after-effects. *J. gen. Psychol.*, 1954, **51**, 291-299.
18. WERTHEIMER, M., & JACKSON, C. W., JR. Figural after-effects, "brain modifiability," and schizophrenia: A further study. *J. gen. Psychol.*, 1957, **57**, 45-54.
19. ZIGLER, E. F., HODGDEN, L., & STEVENSON, H. W. The effect of support on the performance of normal and feeble-minded children. *J. Pers.*, 1958, **26**, 106-122.

Received September 30, 1957.

GENERALIZATION OF CHILDREN'S PREFERENCES AS A FUNCTION OF REINFORCEMENT AND TASK SIMILARITY¹

HAVA BONNÉ GEWIRTZ

University of Maryland

THIS study deals with the effects of certain social reinforcement conditions on the stimulus generalization gradients they produce in children. There have been several recent attempts to apply the concepts of generalization and reinforcement to complex human behavior. Among these, Miller's (7, 8) theoretical models for conflict and displacement phenomena are central. Yet the experiments upon which they are based most often involved infrahuman organisms, physiological reinforcement conditions, and a single dimension of stimulus similarity. And while some studies (4, 12) dealing with complex personality variables have incorporated Miller's conflict theory, generalization in these typically was inferred from assumptions about similarity between objects or situations which could not in themselves be tested because of situational or temporal distance.

In this study, an attempt is made to apply the concepts of generalization and reinforcement to a relatively complex human situation in which the reinforcement conditions are manipulated experimentally. An experiment was designed to investigate children's preferences for a series of problem-solving tasks, as a function of success and failure experiences (positive and negative reinforcers) associated with a training task, and as related to the tasks' degree of similarity to the training task. It was expected that if positive reinforcement followed the solution response to the training task, the differential preference values assigned to the tasks—ordered along the similarity dimension—would represent an approach gradient of preference (i.e., preference for the tasks would increase with their increasing similarity to the training task); and that negative reinforcement applied to the training task would produce an avoidance

gradient (i.e., preference would increase with increasing dissimilarity to the training task). In addition, it was intended to compare the slopes of the two resultant gradients, for a fundamental assumption of Miller's conflict model is that avoidance gradients are steeper in slope than are approach gradients.

Studies of preferences and similar responses (3, 5) have suggested that such behaviors could be acquired according to the laws of conditioning and reinforcement. But it has been demonstrated also (2, 10) that complex social reinforcement experiences such as success and failure may have different implications for subjects with different reinforcement histories relevant to the treatment variable, and that these differences may be reflected in the effectiveness of experimental treatment conditions. This consideration has been taken into account in the design and interpretation of this study.

METHOD

Subjects

One hundred children in the first and second grades of a university laboratory school served as subjects (Ss). They ranged in age from 6-1 to 8-0 years, and their median IQ score was 130.² Children who were considered behavior problems by their teachers and those who were unwilling to participate were excluded.

Materials

A puzzle-solving situation was employed. The material consisted of five Masonite formboard-type puzzles. While all five puzzle frames with their respective covers were of equal size ($11 \times 8\frac{1}{2}$ inches), the diamond-shaped depression within each frame (which constituted the puzzle proper) represented different points along a dimension of shape similarity (Fig. 1). The puzzles at the extremes of the dimension (1 and 5) were used as training puzzles. Each was equipped with two sets of seven triangular plywood pieces to be fitted into the puzzle depression: one was the *easy* set de-

¹ This paper is based on a portion of a doctoral dissertation submitted to the Department of Psychology of the University of Chicago. The writer wishes to express her appreciation and gratitude to Helen L. Koch and to Lyle V. Jones for their guidance in the course of this study.

² IQ scores, most of which were based on the Revised Stanford-Binet, were obtained from the school records. A few scores based on different tests were well above or below the median value, so that it is unlikely that the median would have been altered had all Ss been given the Binet.

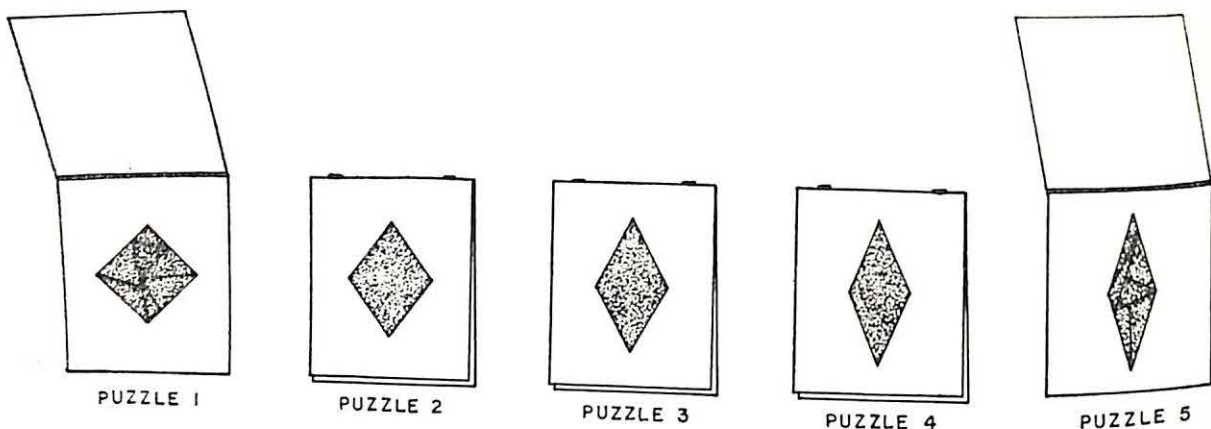


FIG. 1. PUZZLE SIMILARITY DIMENSION

(Puzzles 1 and 5 [the training puzzles] are open and show the "difficult" sets of pieces as placed correctly inside. Puzzles 2, 3, and 4 are closed and their outlines are shown on top of the covers.)

signed to insure successful solution, the other was the *difficult* set designed to produce failure. While both sets were similar in general appearance, the asymmetry of the pieces in the difficult set made solution practically impossible for Ss of the ages employed. The frames and covers of all five puzzles were painted a uniform grey, while the depression in each puzzle as well as all the pieces in the four sets were painted a bright red. In addition, the outline of each puzzle depression was reproduced in red on top of the puzzle's cover. In this way, Ss could observe the shape similarity dimension when the puzzles were covered, without the opportunity to discover clues about the correct placement of the puzzle pieces inside.

Experimental Conditions

Three major conditions were employed: positive reinforcement, negative reinforcement, and control (no reinforcement). The reinforcement conditions represented a combination of three elements. *Positive reinforcement* consisted of (a) objective success in the puzzle-solution attempt, (b) the experimenter's (E's) approval (e.g., "Good," "That was fine!"), and (c) S's winning a material reward or prize (a small plastic trinket of the kind found in gumball vending machines). *Negative reinforcement* consisted of (a) objective failure in the solution attempt, (b) E's reproof (e.g., "That wasn't too good," "Uh-uh"), and (c) the withholding of the material prize. The three elements under each condition were combined in order to maximize the effects of the experimental treatment as well as to minimize individual differences among Ss in their susceptibility to the different kinds of reinforcers involved.

The sample was divided into five groups of 20 Ss each. With the exception of stratification according to sex and grade,³ the assignment of Ss to the five groups

was made at random. These groups did not differ significantly in age or intelligence test scores. Of the five groups, two received positive reinforcement, two received negative reinforcement, and one served as the control. Of each two groups receiving the same reinforcement condition, one was trained on Puzzle 1 and the other on Puzzle 5. The two training puzzles were employed under each reinforcement condition as a crude control for the possibility that puzzles might be differentially attractive due to specific shape characteristics rather than to the effects of the experimental treatment. Thus, in addition to the control group, there were four treatment groups representing the four combinations of reinforcement condition and training puzzle: Positive 1, Negative 1, Positive 5, and Negative 5.

Procedure

The experimental session, in which Ss were seen individually for a period of approximately 20 minutes, was comprised of three consecutive phases: the dimension-training phase, the experimental treatment phase (omitted in the case of control Ss), and the paired-comparison testing phase. The *dimension-training* phase served to familiarize Ss with the puzzle similarity dimension. S was shown the five puzzles, all covered and arranged in a row in a random order. He was asked to rearrange them in terms of their similarity to each other and to verbalize the relevant dimension characteristics. While no S failed to rearrange the puzzles correctly, some were unable to verbalize the basis for their arrangement. In such cases, E provided the relevant information (e.g., "See, they become longer and longer"). This procedure was adopted to insure that Ss were equally aware of the shape similarity dimension.

The *experimental treatment* phase followed when S was led to a separate table and presented with the appropriate training puzzle (either 1 or 5). He was urged to try to solve the puzzle and was told he would win a little prize for each successful solution. At this point S

³ This stratification is ignored in the presentations which follow, because neither age nor sex was found to be related to the experimental results.

was presented with a box containing about 20 prizes and was permitted to select and put aside his three favorites. To convince skeptical *Ss* that solution was possible, *E* preceded the treatment by a brief demonstration, employing the set of pieces on which *S* was to be trained. This was carried out so rapidly that there seemed little danger that *S* would learn the difficult pattern or lose interest in the simple one. Actual treatment began only after this demonstration, and consisted of three trials of puzzle solution. Under the positive reinforcement condition, *S* was provided with the easy set of pieces. After each of the three successful trials he was praised by *E* and allowed to keep one of the three prizes. Under negative reinforcement *S* was provided with the difficult set of pieces. A trial under this condition was defined as an unsuccessful solution attempt, when *S* gave up either spontaneously or following *E*'s suggestion to end it and "start it all over again." Each of the three failures was accompanied by reproof and the removal of one of the previously selected prizes.⁴

The paired-comparison testing phase followed the last training trial, when *S* was asked whether he would like to play some more, but this time with a puzzle of his own choice. The suggestion was welcomed by every *S*. *E* then presented to *S*, successively, the 10 different possible pairs of the five puzzles. During the presentation of each pair, *S* was asked to point to the puzzle with which he would prefer to play. The sequence of pair-presentation was random except that each puzzle appeared an equal number of times in the right and left position in the pair. When the presentation of all 10 pairs was completed, *S* was asked to give the reason for his preference choice ("Why did you pick those?"). Control *Ss*, who were not subjected to the treatment procedure, were presented with the paired-comparison test immediately following the dimension-training phase.⁵

RESULTS

Approach and Avoidance Generalization Gradients

The data were analyzed first by means of the rank analysis method (1, 11), which enabled testing two null hypotheses: (a) puzzle preference equality (i.e., that in each of the

⁴ Five of the positively reinforced *Ss* were noticed to have considerable difficulty in solving the easy puzzle; and three of the negatively reinforced *Ss* almost succeeded in solving the difficult one. These few cases, however, did not change the results and hence are not discussed further.

⁵ Although the experimental session was concluded at this point, *Ss* in the control and negative groups were now permitted to play with the puzzle of their choice, to solve it successfully, and to win the three prizes. This was done in order to temper somewhat the experiences of all *Ss* before their return to their classrooms, and thus to prevent harmful rumors about failure or loss of prizes from reaching children who had not yet had their turn as *Ss*.

five groups there were no significant differences among the preference rankings assigned to the five puzzles); and (b) agreement between groups (i.e., that there were no differences

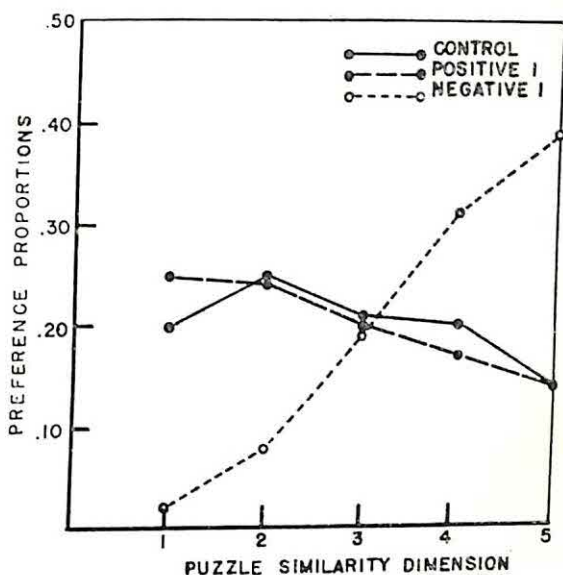


FIG. 2. PREFERENCE GRADIENTS OF CONTROL, POSITIVE 1 AND NEGATIVE 1 GROUPS ($N = 20$ per group)

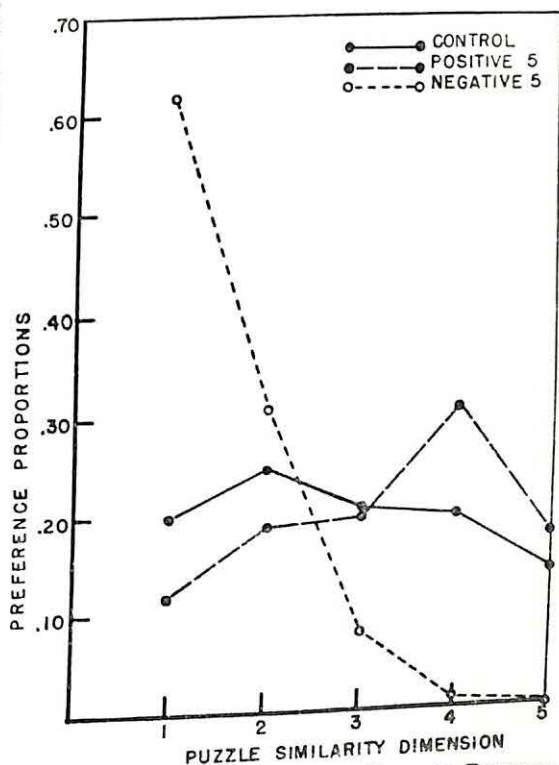


FIG. 3. PREFERENCE GRADIENTS OF CONTROL, POSITIVE 5 AND NEGATIVE 5 GROUPS ($N = 20$ per group)

TABLE 1

CHI SQUARE VALUES FOR THE RANK ANALYSIS TESTS:
PUZZLE PREFERENCE EQUALITY WITHIN EACH
GROUP AND AGREEMENT IN PREFERENCE
PATTERN BETWEEN EACH TREATMENT
GROUP AND THE CONTROL GROUP

Group	Puzzle Preference Equality	Agreement with Control
Control	4.24	—
Positive 1	5.30	1.06
Negative 1	90.31**	59.87**
Positive 5	12.87*	8.75
Negative 5	171.99**	93.41**

Note.—Large χ^2 values (4 df) indicate departure from equality and disagreement.

* $p < .02$.

** $p < .001$.

between each experimental group and the control group in terms of their respective preference patterns).

Figures 2 and 3 show the composite preference gradients (based on preference proportions derived from the rankings) obtained from the groups trained on Puzzle 1 and Puzzle 5, respectively. The control group curve is the same in both figures. Table 1 presents the results of the tests of equality and agreement. As shown, the expectations advanced were fully confirmed only in the case of the two negative groups: each exhibited a reliable avoidance gradient which differed significantly from the preference pattern of the control group. In contrast, the results obtained from the two positive groups were less conclusive: only Positive 5 exhibited significant departure from puzzle preference equality, yet its preference trend was not a simple function of the puzzle similarity dimension; and neither positive group differed significantly from the control group. In addition, response pattern variability within the positive groups was considerably greater than it was within the negative groups, suggesting that a considerable number of Ss were unaffected by the positive treatment condition. This variability was unrelated either to age or sex of Ss, and appeared to be a function of randomly distributed individual differences.

To examine the nature of these differences, it was necessary to determine the number and type of preference response patterns given by individual Ss. Since this information could not be provided by the rank analysis method, another analysis was undertaken. It was based

TABLE 2

NUMBER OF Ss EXHIBITING INDIVIDUAL GRADIENT PATTERNS IN EACH GROUP, AND DIFFERENCE BETWEEN TREATMENT AND CONTROL GROUP FREQUENCIES

Group	Gradient Pattern			Non-gradient Pattern	Difference from Control ^a
	Approach	Avoidance	Total		
Control	3	2	5	15	—
Positive	18	8	26	14	7.02*
Negative	0	30	30	10	11.73*

^a χ^2 corrected for discontinuity, two-tailed test (1 df).

* $p < .01$.

on frequency distributions of Ss classified into two major categories: those exhibiting "gradient" patterns, and those exhibiting "non-gradient" patterns. S was said to exhibit a gradient pattern when the sum of his preference ranks represented the values 4, 5, 6, 7, 8 (or 8, 7, 6, 5, 4), respectively, for the five puzzles arranged in terms of the similarity dimension.⁶ While this particular pattern represented the "perfect" individual gradient possible within the paired-comparison scoring method, slight deviations from this perfect order were included also in the gradient pattern category (e.g., a single reversal in adjacent puzzles; one or two ties on adjacent puzzles provided the order of preference was still maintained in relation to the similarity dimension). All Ss whose sums of ranks did not satisfy these criteria were classified in the nongradient category.

Table 2 presents the frequency distribution of Ss falling within the gradient and nongradient categories, including also classification by gradient direction (i.e., approach or avoidance relative to the training puzzle). Because the different training puzzles did not produce different response gradients in Ss, in this table and in all subsequent ones the data are presented for the combined positive and the combined negative groups, with 40 Ss in each. As shown, the proportion of positively and of negatively reinforced Ss who exhibited gradi-

⁶ These values are based on the paired-comparison scores, where the preferred member in the pair received the rank of 1 and the rejected member received the rank of 2. Each of the five puzzles appeared a total of four times in the ten-pair presentations. Hence, for a single S, the puzzle always preferred would receive the minimal total rank of 4, and the puzzle never preferred would receive the maximal total rank of 8.

ent patterns was significantly greater, in each case, than that proportion among the control Ss. While it is seen that the treatments produced generalization gradients, this analysis reveals also that the positive reinforcement condition produced *two* kinds of gradients: approach and avoidance. This finding could explain the over-all similarity between the control and the positive groups shown in the rank analysis, since the individual gradients, opposite in direction, appear to have cancelled each other in the composite group scores.

Factors Underlying the Effects of Positive Reinforcement

The question that still remained, however, was: why did some of the positively reinforced Ss exhibit avoidance gradients, which had been expected only under negative reinforcement? Since the experiment consisted of a goal-attainment situation, it was assumed that the variations found in Ss' response patterns reflected individual differences in the intensity of their involvement in success, i.e., in their *achievement* motivation. Two classes of information were available in this study for the purpose of testing this assumption: (a) the *reasons* given by Ss for their preference choices, and (b) their *IQ level*. These two variables were taken, each to be a plausible indicator of Ss' strength of achievement motivation (for reasons to be noted subsequently).

Of the various types of reasons given by Ss, two indicated involvement in achievement: preference for easy tasks (e.g., "Because it looks easier", "The others are too hard for me"); and preference for difficult tasks (e.g., "I think it is harder, I like to try it", "Harder is more fun"). These two types of reasons, *Ease* and *Difficulty*, were given by more than 50% of all experimentally reinforced Ss, but only by 10% of control Ss. For IQ level, Ss were divided into two groups: the "High" IQ group consisted of Ss whose scores were above the over-all sample median of 130, and the "Low" IQ group consisted of those with scores below and including that median. It was postulated that the avoidance gradients generated by the positive reinforcement condition were an outcome of strong achievement motivation, and as such would be exhibited more frequently by the more highly achieve-

TABLE 3
NUMBER OF Ss IN THE POSITIVE GROUP CLASSIFIED
IN TERMS OF GRADIENT DIRECTION, IQ LEVEL,
AND TYPE OF REASON

Reason	Gradient Pattern and Direction					
	Approach Gradient		Avoidance Gradient		Non-gradient	
	L	H	L	H	L	H
Ease	10	1	0	3	0	1
Difficulty	0	0	1	4	0	1
Other	5	2	0	0	6	6
Total	15	3	1	7	6	8

Note—"L" and "H" refer to Low and High IQ groups, respectively.

ment-motivated Ss, who were expected to be in the High IQ group and to give *Difficulty* as their reason. Table 3 presents the frequency distribution of the 40 *positively reinforced* Ss in terms of these three variables. When the relationships were tested by means of exact one-tailed tests for fourfold tables (6), the following results were obtained, all at $p < .01$: (a) Ss who exhibited avoidance gradients tended to give *Difficulty* as their reason (5/8), while Ss who exhibited approach gradients tended to give *Ease* as their reason (11/18); (b) Ss who exhibited avoidance gradients were predominantly in the High IQ group (7/8), while Ss who exhibited approach gradients were mostly in the Low IQ group (15/18); (c) when IQ level and type of reason were employed as a joint criterion, it was found that *all* 10 Ss who were simultaneously in the Low IQ and *Ease* categories exhibited approach gradients, while *all* four Ss who were simultaneously in the High IQ and *Difficulty* categories exhibited avoidance gradients. Thus, the postulated relationships between gradient direction, IQ level, and type of reason appear to be supported; and positive reinforcement was shown to generate quite different gradient patterns depending on Ss' strength of achievement motivation.

Relative Steepness of Slope

The examination of the relative steepness of slope of the generalization gradients produced by positive and negative reinforcement involved classification of individual gradient patterns in terms of their degree of *steepness*. The steepest gradient possible under the

TABLE 4

NUMBER OF Ss IN THE TWO REINFORCEMENT GROUPS
EXHIBITING PERFECT AND IMPERFECT APPROACH
AND AVOIDANCE GRADIENTS

Group	Approach Gradient		Avoidance Gradient		Total Gradient	
	Im- Perfect	perfect	Im- Perfect	perfect	Perfect	Im- perfect
Negative	0	0	22	8	22	8
Positive	8	10	2	6	10	16

method employed was the "perfect" gradient, described earlier. All other gradient patterns were classified as "imperfect." Table 4 presents the distribution of Ss in the two reinforcement groups in terms of this steepness criterion, and also according to gradient direction. Two-tailed tests (χ^2 corrected for discontinuity, 1 *df*) were used. The first question addressed itself to the effects of each treatment condition on steepness, ignoring gradient direction. It was found that the proportion of steep gradients that were obtained after negative reinforcement (22/30) was significantly greater ($p < .02$) than the proportion obtained after positive reinforcement (10/26). When gradient direction was taken into account, however, it was found that *perfect* avoidance gradients were produced more frequently ($p < .01$) in the negative group (22/30) than in the positive group (2/8); but that no differences existed in the positive group between the proportion of its perfect avoidance gradients (2/8) and the proportion of its perfect approach gradients (8/18). These results indicate that, regardless of gradient direction, negative reinforcement produced more and steeper individual gradients than did the positive reinforcement condition.

DISCUSSION

Achievement Motivation and Preference Gradients

It has been suggested (e.g., 2) that children's preferences for difficult goals are a function of strong achievement emphasis during socialization. Similarly, when, following success, S verbalized preference for a difficult task in the present study, it was taken to indicate that he was more motivated to demonstrate outstanding achievement than was another S who

verbalized preference for the easy task on which success had already been experienced. The use of IQ level as the second index of achievement motivation was based on the notion that ability is a significant determinant of the *outcome* of past encounters with difficult tasks. Highly intelligent children are likely to be rewarded for attempts to excel on such tasks, either because of actual success, due to superior ability, or because parental approval is often contingent upon their accomplishing more than "just doing what everybody else can do." Hence, the highly able child should be more likely to acquire strong achievement motivation than the less able child. Another possible approach could be based on the notion that high IQ scores are, in part, an outcome of strong achievement motivation. The child who persists in his problem-solving efforts and refuses to give up easily is more likely to attain a higher score than one who is indifferent or gives up at first sight of difficulty.

These assumed relationships, regardless of their causal direction, appear to be supported by the present experimental results. Mastery of the difficult task appeared to have a stronger reinforcing value to the highly achievement-motivated S than repeated success on the easy one. Hence, it would seem more correct to conclude that rather than generating approach and avoidance gradients, the positive reinforcement condition produced *two kinds of approach gradients*: one by the less achievement-motivated Ss, directed towards the goal of easy and safe success (the approach gradient); and the other by the highly achievement-motivated Ss directed towards the goal of outstanding achievement in terms of solving the most difficult appearing task (the avoidance-like gradient). This conclusion seems supported also by the finding that these two kinds of gradients resembled each other in terms of the steepness criterion, regardless of direction.

Negative Reinforcement Generates the Steeper Gradient

The limitations inherent in the paired-comparison method⁷ restrict somewhat statements

⁷ Because the lack of independent response measures for each of the compared items, the effects of reinforcement on the response to each of the five puzzles is confounded with its effects on the discrimination of the training puzzle from all others.

about the specific shape of the generalization gradients obtained. Nevertheless, the marked differences between the effects of the two treatment conditions suggest certain speculative interpretations. While the distinction between primary and acquired drives, as suggested by Miller and Murray (9), does not fit the case of the present study, an analogous differentiation between "internal" and "external" sources of motivation for the response might be applied here. The fact that no individual differences could be detected in the uniform response pattern exhibited by the negative group suggests that such differences—though theoretically present—were overshadowed by the powerful and unambiguous negative treatment condition. The avoidance response appears to have been contingent predominantly upon the external stimulus situation which, when changed, was readily discriminated and produced a decrease in response strength resulting in a steep generalization gradient. On the other hand, the variability in the positive group's response pattern indicated that this treatment condition was responded to by Ss primarily according to their individual aspirations and achievement needs. The "need to succeed" (being different perhaps for each S) was relatively more constant and more independent of the external stimulus situation. Hence, changes in that situation did not bring about a marked, or at least uniform, reduction in response strength, making for flatter generalization gradients than those produced under the negative reinforcement condition.

SUMMARY

An experiment was designed to study children's preferences for a series of problem-solving tasks as a function of the particular reinforcement condition associated with a training task, and the degree of similarity of each task to the training task. One hundred first- and second-graders served as Ss. Five formboard-type puzzles were constructed such that their shapes constituted a similarity dimension. The two puzzles at the extremes of this dimension served as training puzzles, each for one-half of a group receiving the same reinforcement. Forty Ss received positive reinforcement, consisting of successful puzzle solution, praise, and a material reward; 40 Ss

received negative reinforcement, consisting of failure, reproof, and withdrawal of the reward; and 20 Ss were not reinforced, serving as a control group. Ss' differential preferences for the five puzzles were obtained through subsequent paired-comparison presentations.

The results were: 1. Following reinforcement, the differential preference rankings for the puzzles (arranged in terms of the similarity dimension) represented generalization gradients of preference. But while negative reinforcement uniformly generated avoidance gradients, positive reinforcement produced in some Ss approach gradients and in others avoidance gradients. 2. Those positively reinforced Ss who exhibited avoidance gradients tended to have superior IQ, and to verbalize preference for difficult tasks. It was postulated that these Ss were more achievement-motivated than those who exhibited approach gradients, and that their response pattern actually represented approach towards a more challenging task. 3. When the slopes of the gradients were compared, it was found that negative reinforcement generated a larger proportion of steeper individual gradients than did positive reinforcement. It was suggested that this finding could be related to the distinction between external and internal factors which determined the response.

REFERENCES

1. BRADLEY, R. A., & TERRY, M. E. Rank analysis of incomplete block designs I: The method of paired comparisons. *Biometrika*, 1952, **39**, 324-345.
2. CHILD, I. L. Children's preferences for goals easy or difficult to attain. *Psychol. Monogr.*, 1946, **60**, No. 4 (Whole No. 280).
3. GEBHARDT, MILDRED E. The effect of success and failure upon the attractiveness of activities as a function of experience, expectation and need. *J. exp. Psychol.*, 1948, **38**, 371-388.
4. HOLLENBERG, ELEANOR, & SPERRY, MARGARET. Some antecedents of aggression and effects of frustration on doll play. *Personality*, 1951, **1**, 32-43.
5. HUNT, D. E. Change in goal-object preferences as a function of expectancy for social reinforcement. *J. abnorm. soc. Psychol.*, 1955, **50**, 372-377.
6. LATCHA, R. Tests of significance in a 2×2 contingency table: Extension of Finney's table. *Biometrika*, 1953, **40**, 74-88.
7. MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 431-465.

8. MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155-178.
9. MILLER, N. E., & MURRAY, E. J. Displacement and conflict: Learnable drives as a basis for the steeper gradient of avoidance than of approach. *J. exp. Psychol.*, 1952, **43**, 227-231.
10. SEARS, PAULINE S. Level of aspiration in academically successful and unsuccessful children. *J. soc. Psychol.*, 1940, **35**, 498-536.
11. TERRY, M. E., BRADLEY, R. A., & DAVIS, L. L. New designs and techniques for organoleptic testing. *Food Tech.*, 1952, **7**, 250-254.
12. WHITING, J. W. M., & CHILD, I. L. *Child training and personality: A cross-cultural study*. New Haven: Yale Univer. Press, 1953.

Received October 7, 1957.

SOCIAL INFLUENCE ON OPINIONS AND THE COMMUNICATION OF RELATED CONTENT¹

BERTRAM H. RAVEN

University of California, Los Angeles

THE past decade has witnessed a phenomenal increase in theories and investigations of the group pressures and interactions which results from discrepancies in opinion. Such differences, Festinger and his co-workers observed (13, 14, 15, 16, 32), result in three distinct manifestations of pressure toward uniformity, each of which operates to reduce this discrepancy: a pressure on the individual to alter his own opinion so as to bring it in line with his fellows; a pressure on the individual or group to attempt to influence discrepant individuals so as to bring their opinions in line; a pressure on the group or individuals to redefine the boundaries of their group, increasing uniformity by rejecting discrepant individuals. The results of these investigations have contributed considerably to our understanding of social influence. They have also indicated a number of problems which need additional investigation.

First, we still have much to learn about the factors that increase pressures toward uniformity in the group and of their effects on the individual in relation to his group. Group pressures have been shown to increase with cohesiveness (1, 16, 32), homogeneity (17, 19), and clarity of group goals and procedures (31), and to decrease when opinions are anchored in other groups (20, 25). Though rejection by the group has been shown to be a manifestation of pressures toward uniformity (32), there has been little investigation of the effects of fear of rejection on pressures toward uniformity.²

¹ This study was conducted under contract with the Office of Naval Research while the author was a member of the staff of the Research Center for Group Dynamics, University of Michigan. The more complete dissertation upon which this report is based (29) was submitted to the faculty of the University of Michigan in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to express his appreciation to John R. P. French for his patient direction, to Leon Festinger for the initial inspiration, to Irwin Goffman, and Stanley Thorley for their invaluable assistance in the analysis of the data.

² Since this investigation, there have been several reports of the effects of fear of rejection (9, 22, 24).

Secondly, there is still the question of the relationship between the social influences on opinion and influences on the content, perceived or communicated, which bears on that opinion. Duncker (10) may have been one of the first to observe that "prestige suggestion" might be the result of new information and new perceptions about the object of opinion, rather than blind conformity. Or as Asch (1) states it, prestige suggestion might be "a change in the object of judgment, rather than in the judgment of the object." Asch, Block, and Hertzman (3) demonstrated experimentally that subjects who altered their evaluation of the profession of "politics" to conform to a majority evaluation also altered their perception of the connotation of the word "politics." However, whether alteration of perception of the object of judgment preceded the alteration of the judgment, as the experimenters suggested, or whether the changes in perception rather occurred *after* the change in judgment is not clear. The distinction between social influence on opinion and on content bearing on the opinion was, nevertheless, clearly established. Further investigation to determine relationship between these two types of pressure seems in order.

Thirdly, the distinction has been made between that change in the individual's opinion which is evident to the group or "public" and that which is "private" (33). More recently there have been several attempts to delimit the conditions that result in private acceptance of a change in opinion as compared with those that result in public conformity without private acceptance (8, 14, 18, 25, 26, 30). We would further expect that even though the person's opinion may remain private, insofar as his group is concerned, there will be pressures to change toward the group, and these will be accelerated if he must communicate content that is related to the opinion.

It is our purpose here to investigate a situation where the individual finds his own opinion discrepant from the clearly defined

norms of his group, but where he may keep his *opinion* private. We would vary the group pressures upon the individual by allowing some groups the opportunity to reject the individual for nonconformity. In addition, we wished to vary the extent to which his communications regarding the opinion would remain private. Where he must communicate to his group material relating to the discrepant opinion, pressure to change should be increased. Furthermore, the group pressures would operate to distort the content relating to the opinion even though the opinion itself would remain private. These pressures would be greater with respect to content that is to be communicated to the group than upon content that the individual perceives. Pressures toward uniformity operating on opinion, on content communicated, and on content perceived should be increased when there is possibility of rejection for nonconformity. It was to examine such hypotheses that this experiment was designed.

THE EXPERIMENT

Procedure

There were from 10 to 14 undergraduates in each experimental group, a total of 344 subjects (Ss) in all. Each group was homogeneous with respect to sex and class status. The cohesiveness of each group was heightened by a procedure utilized by Back (4), by using only those Ss who had expressed an interest in the subject matter proposed for discussion—"human relations and social problems," and by a short group discussion.

Each S was then given a juvenile case study to read. "The Case of Johnny Sandron" was adapted from an actual case study (12). Johnny had been arrested for the robbery and murder of an elderly lady. Interviews with Johnny, with his mother, and with his teacher were presented, each providing contradictory information about Johnny and his crime. When the initial reading had been completed, Ss were asked to indicate their opinion on a seven-point scale wherein they evaluated

the extent to which they felt that Johnny was personally responsible for his crime:

1. Johnny had many decent influences of which he could have taken advantage. The conditions under which he lived were friendly enough so that they can hardly be held responsible for his misdeeds. *The blame for his crime must be placed entirely upon his shoulders.* . . . 4. Johnny had both helpful and disturbing influences. *The blame for his crime must be placed equally both on Johnny and on conditions in which he lived.* . . . 7. Considering the terrible conditions under which Johnny lived, it would seem almost a miracle that he didn't come out worse than he did. In spite of everything, Johnny's behavior was still essentially good. *Not he but his environment is entirely to blame for his misdeeds.*

It was emphasized to the Ss that their opinion was stated privately, to the E only, and would not be communicated to fellow group members.

The next operation was designed to make as many Ss as possible feel that they were deviates from a well-defined group norm:

You are probably curious about how others in this group think about the extent to which Johnny is responsible for his crime. Therefore I shall place on the blackboard the number of people who have chosen each position. I shall indicate this with Xs so that it will not be necessary to identify each person specifically.

Most Ss tended to choose opinions along the environmental responsibility end of the scale, and a false consensus was presented which indicated that most Ss occupied an opinion at the "personal responsibility" end of the scale (see Fig. 1). Thus, Ss who occupied lone positions at 5, 6, or 7 saw most of their group members at 2 or 3. Ss at 5, 6, or 7 will henceforth be called the *deviates*; those who initially held Positions 2, 3, or 4, will be called *modes*. The reaction to the fictitious consensus was often one of shocked surprise. However, only a few Ss were suspicious of the manipulation, and these were readily detected and eliminated from the analysis of the data.

Eventually, Ss were told that they would have to write a group report of the case, much as social workers must do. This would be a report upon which all must agree completely. A prize would be awarded to the group which submitted the best report. With this in mind, Ss were now asked to reread the case study, and a second statement of opinion was solicited. Again, the opinion statement was private, not to be seen by fellow Ss, and it was possible to compare the new opinion statement with the initial one to get a measure of the effects of group pressures on opinion.

As further aid in writing their *group* report, Ss were next asked to write individually a description of the case as they saw it. A rough outline was provided by the E. The description was solicited by the E so that it might later be coded for group influences on content related to the opinion in question.

Following the individual descriptions of the case, Ss were asked to state their private opinion for the third time. This was followed by a questionnaire. Then, instead of the group discussion and group report, a complete explanation of the research project was presented by the E, and the session was concluded.

Opinion Scale Position	Actual Consensus	Presented Consensus
1		
2		XXXXXXX
3	A	XX
4	BC	X
5	DEFGH	X
6	IJKL	X
7	M	X

FIG. 1. A TYPICAL OPINION CONSENSUS FOR A GROUP OF THIRTEEN AND THE FICTITIOUS CONSENSUS PRESENTED TO THAT GROUP

The Public-Private Variation

Though it was stressed to all Ss that their own opinions need never be stated publicly to the group, this did not hold for the descriptions of the case study which they wrote individually. In the *Public* condition, Ss were told that their individual descriptions would be passed around so that all other group members could see them. To facilitate such distribution, carbon copies were made of the individual descriptions. In the *Private* condition, the individual descriptions were not to be seen by anyone other than the *E*. The purpose of the *Private* descriptions was presumably to aid Ss in formulating their own ideas. Thus in the *Public* condition, Ss were forced to communicate opinion-relevant content to members of their group, whereas such communication was not forced on Ss in the *Private* condition. Furthermore, the content of the *Public* descriptions would indicate group influences on communication. The content of *Private* descriptions could be considered as indicating primarily group influences on perception or cognition.

The Variation of Possibility of Rejection

In order to observe the effects of differing degrees of pressure toward uniformity, we also varied possibility of rejection. Ss in the *Rejection* conditions were told that, since it was especially important that members of the group get along well together, we were offering them an opportunity to reorganize the group so as to exclude those members with whom they could not get along well. After they had had an opportunity to discuss part of the case study, ballots would presumably be passed out on which they could indicate their candidates for rejection. The three individuals with the most rejection votes would then leave to take part in another research project. The *Nonrejection* Ss were told that all of them would remain for the entire group session.

Both the *Rejection* and *Public-Private* variations were introduced at the very beginning of the experiment and were reinforced at various times throughout the session. Thus we had four experimental conditions: *Public Rejection*, *Public Nonrejection*, *Private Rejection*, *Private Nonrejection*. Within each of these conditions, Ss were divided into *modes* and *deviates*

TABLE 1
DISTRIBUTION OF SUBJECTS IN EXPERIMENTAL CONDITIONS

	Public		Private	
	Modes	Deviates	Modes	Deviates
Rejection	24	74	18	76
Nonrejection	11	67	17	57

viates according to whether they held initial opinions which were in agreement with the initial group norm. The number in each condition is shown in Table 1.

Data Collection

Three types of data were collected: opinion statements, questionnaire responses, and content analysis of case descriptions.

1. *Data on change of opinion.* Three private statements, regarding the extent to which Johnny was responsible for his crime, were obtained from the Ss. One was obtained after the first cursory reading of the case study. The second statement was made after the fictitious consensus had been presented and after a rereading of the case study. The third statement was made after the individual report had been written. Change of opinion, and direction of change (toward or away from the group norm), could be measured from a comparison of these opinion statements.

2. *Questionnaire data.* The questionnaire administered at the end of the experiment was constructed primarily to test the effectiveness of the experimental manipulations.

3. *Content analysis of individual descriptions.* To gain information regarding the Ss' tendency to select or distort content in response to group pressures, the content of the individual descriptions was analyzed. Each description was broken down into a number of meaning units or items. Each item was now coded in terms of whether it was "favorable" (supporting the end of the opinion scale which was favorable to Johnny), "unfavorable" (supporting a position unfavorable to Johnny), or "neutral" (not clearly supporting either end of the scale). Thus "favorable" items tended to support the position actually chosen by most Ss; "unfavorable" items tended to support the presumed group norm. After adding the number of items in each category for the individual

description, a "coefficient of imbalance" (CI) was calculated, subtracting the percentage of unfavorable responses from the percentage of favorable responses (23). An arcsine transformation was used to make the distribution of scores approach normality.³ Thus, the greater the CI, the more positive the content and, among deviates, the greater number of items supporting the S's own initial opinion, favorable to the delinquent. The less positive, or more negative, the CI, the more the deviate shows evidence of selecting or interpreting the content so as to support the supposed group norm.

The Effectiveness of the Experimental Manipulations

The effectiveness of the Public-Private manipulation was tested with responses on the final questionnaire. One question asked, "How much did you feel that your opinion (as to whether Johnny or his environment was responsible) would affect the degree to which the others in the group would prefer to have you continue working with them on the case study?" with a six-alternative scale. The deviate Ss in Public conditions were much more likely to see a relationship between opinion and rejection than those in the Private conditions ($p = .04$, by chi square), indicating that Ss did indeed feel that their opinions would be more obvious if their descriptions were seen by others. However, another question, "What possibility is there that the other members of the group will find out how you personally feel about this case study?" showed no significant differences, perhaps because of the ambiguity of the phrase

³ The precise formula for the coefficient of imbalance was:

$$CI = \arcsin \frac{f}{f + u + n} - \arcsin \frac{u}{f + u + n},$$

where f = number of favorable items, u = number of unfavorable items, n = number of neutral items. The possible range of scores was from + 90.00 to -90.00. For all Ss, the mean CI was + 1.03, the SD was 19.3. A reliability check between two independent judges, based on 33 cases, and a total of 777 items, showed agreement on 79.8%, with only 2.9% in which one judge coded an item as positive and the other as negative. The correlation between the 33 pairs of CIs was .84. The validity of the CI was demonstrated by a strong relationship between CI and initial opinion, significant at the .001 level by chi square.

"how you personally feel" ($p = .20$, by chi square).

To test the realism of the rejection manipulation, we asked "How likely is it that the others in the group would prefer that you not continue working with them?" There were sharp differences between Rejection and Nonrejection Ss in their choice of the six alternatives with respect to this question, Rejection Ss perceiving a much greater possibility of rejection ($p = .001$, by chi square). Deviate Ss in the Rejection conditions were also more likely to feel that their opinion would be the basis for rejection than did those in Nonrejection conditions, their responses to the question above also being statistically significant ($p = .001$, chi square). The interaction effect between rejection and publicity was also significant ($p = .05$, $2 \times 2 \times 2$ chi square). Deviates in the Public Rejection condition were most likely to feel that their opinions would influence their acceptability; deviates in the Private Nonrejection condition were least likely to see this relationship.

The false consensus was also effective in giving deviate Ss the impression that they held opinions which were sharply discrepant from the group, even though their opinions in actuality were in line with the majority. The responses to the question "How much did your first opinion differ from that of the group as a whole?" showed sharp differences between deviates and modes ($p = .001$, chi square). Deviates were also more likely to associate their opinions with rejection than were modes ($p = .002$, chi square).

In general, our manipulations seemed effective, though we could have hoped for more clear evidence to this effect with respect to the Public-Private variation.

HYPOTHESES AND RESULTS

We had thus organized a number of cohesive groups, each with a clearly defined group norm from which a number of the Ss felt that they deviated. There was pressure on the group to achieve uniformity. Ss wrote individual descriptions of a case study which in some groups were to be made public, in others were to remain private. In all groups, however, the individual's *opinions* were kept private. In some groups, the Ss were convinced that deviates might be rejected for noncon-

formity; in others no such possibility of rejection existed. The experiment was designed to test two sets of hypotheses, some with respect to pressures on opinion, others dealing with pressures on content:

Group Pressures on Opinion

Given a situation such that (a) an individual is a member of a group toward which he is attracted, (b) he deviates from a well-defined norm of that group, (c) the individual has no intense involvement in his initial opinion, (d) there is pressure on the group to achieve uniformity with respect to that opinion, (e) the individual does not perceive that he can readily achieve uniformity by influencing others:

Hypothesis 1: There will be pressures on the individual to change his opinion toward the group norm.

This hypothesis is consistent with findings by Asch (2), Sherif (34), Festinger (13, 15), and others (4, 8, 16, 19, 20, 26). We note, however, that this pressure to change is predicted even though the opinion is to remain private. Specifically, we predict that the deviates would tend to change their opinions more than the modes, such changes being in the direction of the supposed group norm. Table 2 shows that this hypothesis was clearly supported, as evidenced in comparison of initial and final statements of opinion. Whereas modes tended to shift seldom, and then in either direction, a very large percentage of the deviates changed toward the group norm.

Hypothesis 2: The more the individual must communicate regarding the object of opinion, the greater will be the pressure to change his opinion toward the group norm.

This would occur even if the opinion itself were to remain private. The communication will of itself make the individual more aware of his discrepancy from the group norm, and emphasize the importance of conformity. In addition, we suspect that the pressures which affect communicated content operate upon perceived content, eventually effecting change in opinion. We shall discuss this further below. As we can see in Table 3, the Ss in the

TABLE 2
NUMBERS OF MODES AND DEVIATES WHO CHANGED THEIR OPINIONS^{a, b}

	Initial Position	Changed toward Position 2	Did not Change	Changed toward Position 7
Mode	2, 3, or 4	12	46	12
Deviate	5, 6, or 7	89	168	16

^a Supposed group norm was at Positions 2 and 3. Comparison is between initial and final statement of opinion.

^b Deviates, at Positions 5, 6, and 7, were much more likely to change toward the opposite extreme than were the modes, who supposedly had group support—89 of the 343 deviates made such a change, compared to 12 of the 70 modes. This difference in proportions is significant at well below the .001 level of confidence. Comparable figures on the basis of the second statement of opinion were 85 for deviates and 8 for modes, of the same total.

TABLE 3
PROPORTION OF DEVIATES WHO CHANGED TOWARD THE GROUP NORM^a

	Public	Private	Combined
Rejection	.41	.30	.35
Nonrejection	.36	.17	.29
Combined	.39	.26	

^a These figures are based upon comparison between first and final opinion statement. Comparable proportions for change at the second statement are: Public Rejection, .36; Public Nonrejection, .34; Private Rejection, .29; Private Nonrejection, .23. These proportions are not significantly different from those shown above.

Public conditions were more likely to change than those in the Private conditions, the combined proportions being significantly different at the .02 point (*t* test of difference in proportions). This difference was especially marked among the Nonrejection Ss ($p = .03$). The lesser difference among Rejection Ss may be due to the fact that, with 41% of the Ss changing in the Public Rejection condition, we had reached some upper limit with respect to willingness to accept group influence. The hypothesis can be said to be substantially supported.

Hypothesis 3: The greater the possibility of rejection for nonconformity, the greater the pressure to change toward the group norm.

If fear of disapproval can serve as a source of pressure toward uniformity, possibility of overt rejection should accentuate such pressure. Again, the prediction was made for deviates even though they need not publicly state their opinions. Dittes and Kelley (9) have since found additional evidence to support such an hypothesis.

Though the differences were in the predicted direction, rejection alone seemed to have no significant effect upon change in opinion, the combined proportions, comparing Rejection and Nonrejection deviates, being different at the .13 level of significance (t test). However, pressure to change opinion was least for deviates who did not fear rejection and also had assurance that their descriptions would not be made public. The difference between the Private Nonrejection condition and the other conditions was significant at the .02 level. Only 17% of the deviates in the Private Nonrejection condition changed their opinions toward the group norm, compared to 41% for the Public Rejection deviates ($p = .01$, by t test). Obviously, the effects of rejection must be considered in conjunction with publicity.

Pressure to Change and Opinion Scale Position

Though no prior hypotheses were formulated with respect to the relationship between change in opinion and the extremity of opinion, such a relationship did emerge in the analysis, and it will assume some importance in the discussion that follows. Studies of attitudes have presented evidence that seems to relate intensity of an attitude with the extremity of that attitude. Cantril (6) plotted extremity of attitude toward government regulation of business against intensity and found a U-shaped curve. Guttman and Suchman (21) have adopted intensity as a means of establishing the zero point of an attitude. A similar relationship was found by Kelley and Volkart (25). We might expect that, in this experiment, deviates who held Position 5 in the first opinion census would feel less strongly about their opinion than those at Position 6 or 7, and should therefore be more susceptible to pressures to change. In fact, 32% of the Ss at the extreme positions changed, while 39% of those at Position 5

changed their opinions toward the norm. Though these proportions do not differ significantly, one must also consider the fact that Ss at Position 6 or 7 had a greater number of positions toward which they could change. The effects of extremity on change of opinion are even more evident in the amount of change (Table 4). Of those deviates who did change, we find that those at extreme initial positions were likely to change only one step, while those at Position 5 were just as likely to change two and three steps. The difference here is significant ($p = .002$, by chi square).

Group Pressures on Content

Hypothesis 4: Group pressures to select and distort content regarding the object of opinion will be greater with respect to content that is to be communicated to the group than on content that is perceived by the individual.

Assuming that fear of disapproval or rejection by the group is prominent, the individual in his communication would be expected to select those items from the content he perceived that support the group norm, to avoid items of content that oppose the group, and to distort ambiguous items so that they seem to support the group norm. Thus the content communicated would indicate additional selection and distortion over that perceived. This is, again, consistent with studies by Schanck (33) and others who have noted differences between publicly expressed opinions and those which are privately held, but we go even further and suggest that even though the opinion itself may remain private, communication regarding the object of opinion will be distorted.

Hypothesis 5: Group pressure to select and distort communicated and perceived content so as to support the group norm will be greater when there is possibility of rejection for non-conformity.

TABLE 4
NUMBER OF DEVIATES CHANGING ONE, TWO, THREE,
AND FOUR STEPS TOWARD THE NORM

Initial Position	Number of positions changed			
	1	2	3	4
5	17	16	1	0
6	25	3	1	1
7	5	0	0	0

Under the conditions outlined, more overt pressure from the group might influence both communicated and perceived content, even though the opinion itself is private. The group effects on communication, outlined above, should be increased when possibility of rejection is more salient. Furthermore, we expected this increased effect on the individual's perceptions as well. Let us assume that the individual ex-

amines the object of opinion with a view toward communicating acceptable items to his group. Eventually, this need for acceptance and approval would result in selective and distorted perceptions, just as other needs have been shown to influence perception (5, 28). Accordingly, increased group pressure would influence both perceived and communicated content.

From Hypothesis 4 we should predict that deviates in the Public conditions should show more selection and distortion of content toward the group norm (less positive CIs) than those in the Private conditions. From Hypothesis 5, deviates in the Rejection conditions should have less positive CIs than those in the Nonrejection conditions. CIs should be most positive in the Private Nonrejection condition and least positive in the Public Rejection condition. Even for those Ss who did not change their opinions, group pressures should affect communication and perception.

The first test of this hypothesis comes then from an examination of the CIs for those deviates who did not change their opinions toward the group norm (Table 5). It is evident that the hypotheses were not supported for these Ss. The Public Nonrejection condition showed a less positive CI than the Private Nonrejection condition ($p = .02$ by t test) but even this difference is questionable since the Public Nonrejection CIs were also more variable ($p = .01$, by F test).

There is some question whether the deviates who did not change their opinions provide an adequate test of the hypotheses. As the result of differing pressures on opinions, the opinion distributions of nonchanging deviates at the time of the writing of the descriptions differed for the various experimental conditions. We might therefore get a more adequate test from those Ss who held the same position at the time that they wrote descriptions. Thus, we see in Table 6 the CIs of all deviates who held Position 5 according to their final opinion statement. By our previous analysis, these Ss should be relatively less involved in their own opinion than those who held Position 6 and 7. The differences here are all in the predicted direction. The difference between the Public and Private deviates is significant at the .05 level (t test), supporting Hypothesis 4. This is particularly true for Ss who did not fear rejection ($p = .03$, by t test), but less true for

TABLE 5

MEAN COEFFICIENTS OF IMBALANCE FOR DEVIATES WHO DID NOT CHANGE THEIR OPINIONS^{a, b}

	Public	Private
Rejection	8.14 (38)	8.69 (42)
Nonrejection	.69 (36)	9.21 (36)

^a Difference between Public Nonrejection and Private Nonrejection significant at .02 by t test, but Public Nonrejection variance was also greater ($p = .01$, F test). Other differences were not significant.

^b The figures in parentheses represent the number of Ss in each category. A more positive CI represents a description containing a greater amount of content supporting the "deviate" end of the opinion scale.

TABLE 6

MEAN COEFFICIENTS OF IMBALANCE FOR DEVIATES WHO HELD POSITION 5 AT FINAL OPINION STATEMENT^{a, b}

	Public	Private	Combined
Rejection	.85 (28)	4.73 (31)	2.89 (59)
Nonrejection	1.42 (28)	12.30 (17)	5.53 (45)
Combined	1.13 (56)	7.42 (48)	

^a The mean coefficient of imbalance of the Private Nonrejection condition is significantly greater than that of the Public Rejection condition and the Public Nonrejection condition (in both cases, $p = .03$ by t test). The difference between the combined Public and combined Private conditions is significant at the .05 level. No other differences are significant at the .05 level.

^b The figures in parentheses represent the number of Ss in each category. A more positive CI represents a description containing a greater amount of content supporting the "deviate" end of the opinion scale.

those in the Rejection condition. Again, we might surmise that some upper limit had been reached by Ss in the Public Rejection condition, reducing the difference between that condition and others. Thus Hypothesis 4 is supported for those Ss who do not have strong involvement in their opinion, and choose a less extreme position. The differences for Ss who held Positions 6 and 7 at the final census were not significant.

The differences between the Rejection and Nonrejection deviates at Position 5 were not significant ($p = .24$). The combined effect of publicity and possibility of rejection is clear, however, the difference between the Public Rejection and Private Nonrejection CIs being significant at the .03 level by t test (Table 6). Thus the data on pressure to distort content parallels that on pressure to change opinion. Rejection, in and of itself, does not seem to significantly affect distortion of content. However, deviates who are assured of safety from

rejection, and of privacy in their descriptions of the case study, experience least pressure to distort content toward the group norm, and are significantly more likely to write content supporting their own opinions.

Content and Opinion Change

Implicit in much of the foregoing discussion is the assumption that group pressures might influence opinion by first affecting the communicated and perceived content that bears on it. After the perceived content has been influenced, the individual will alter his opinion so as to make it consonant with the object as he now sees it. There are some data to support such an assumption. Seventeen Ss in all changed their opinions toward the group norm between their second and third statements—that is, during or shortly after the writing of their individual descriptions. We compared the mean CIs of these people with those of Ss at the same initial positions who had not changed their opinions. As indicated in Table 7, 15 of these 17 Ss had CIs that were more negative (more supportive of the group norm) than those of Ss at equivalent positions who had not changed. The probability of obtaining that large a proportion by chance is significant at the .002 level (binomial expansion). It seems evident that there was in fact a tendency for these Ss to bring their opinions in line with the content that they perceived or communicated.

TABLE 7

NUMBER OF DEVIATES CHANGING THEIR OPINIONS WHILE WRITING DESCRIPTION, SHOWING TENDENCY TO MAKE OPINION CONSONANT WITH CONTENT

Position Before Changing	Frequency of CIs of Changing Deviates, Relative to Mean CIs of Nonchanging Deviates ^{a, b}		
	More negative	More positive	Total
5	11	2	13
6	4	0	4
Total	15	2	17

^a As in previous tables, a more negative CI contains a greater amount of content supporting the supposed group norm; a description with a more positive CI contains more content supporting an opinion at the "deviate" end of the opinion scale.

^b Comparisons are based on changes between the second and third opinion statement. Mean CIs for Ss at Position 5 who *did not* change while writing their descriptions was 4.27; 11 of the 13 Ss who *changed* while writing had CIs which were less than 4.27. Non-changing Ss at Position 6 had an average CI of 14.52; all 4 deviates who *changed* opinions between the second and third statements had less positive CIs. The probability of 15 of 17 deviates who changed having more negative CIs is .002. No Ss at Position 7 changed between the second and third statement.

DISCUSSION

One of the original purposes of this study was to distinguish experimentally between pressures to change opinions and pressures to distort content related to the object of opinion. It is clear that group pressures can operate at both levels. The hypotheses regarding distortion received less strong support than those about group pressures on opinion, perhaps in part because the measurement of opinion change is more direct than measurement of content. Also the forces affecting content related to an opinion would seem more diffuse than those that affect opinions. There are indications, however, that the effects of group influence on opinions are more subtle than generally supposed.

The evidence from this study strongly suggests that opinions can be influenced through first influencing content. This is the position that was taken by Asch, Duncker, and others (1, 2, 10). However, the theory posed here is that such influence may not occur merely through the acquisition of additional information. If an individual finds himself in a situation such that he is attracted toward a group, fears rejection from that group—either overt or through ridicule—and finds that his opinion is sharply different from that of the group, it seems that he might first inhibit his expression of specific content that is not supportive of the group norm and might prove disturbing to the group, or distort his communication of ambiguous material so that it supports the group norm. He next may begin to look for content that he can communicate—the influence thus being on his perceptions of the content related to group opinion. Studies of the "influence of needs on perception" (5) suggest that such influence need not be conscious. Finally, the individual actually is aware of more content supporting the group norm than supporting his original opinion, and, in this sense, he has an "objective" basis for changing; the total data available to him support the group opinion rather than his original position. In effect, he changes his opinion so as to make it consonant with his phenomenological data. Such a theory may help to explain the situation that occurs when white Northern students in a Southern university gradually adopt the racial attitudes of

their classmates. By the very fact that he is attending a Southern university, the student indicates some attraction toward his fellow students and a desire for acceptance by them. He is initially aware of the discrepancy in attitudes towards Negroes, and may inhibit the expression of information available to him that would place Negroes in a favorable light. He next begins to *look* for things that he can communicate, and will remember items that tend to support the Southern white point of view, distorting or overlooking items that he cannot comfortably communicate. We might expect that an informational questionnaire about Negroes would show that the Northern student has accentuated those characteristics that support the Southern white stereotype well before he actually adopts the Southern attitude. Finally, we might find that the Northern student has accepted the attitude itself, bringing his attitude into consonance with his perceptions. This is, of course, only one path that pressures toward uniformity might take. Nor can we claim support for this theory merely on the basis of evidence in this study. Additional investigation would seem well in order.

SUMMARY

In this study, we attempted to determine the effects of group pressures on opinion, where opinion was to remain private. We were also interested in group influences on content related to the opinion, both as perceived and as communicated. We wished to determine how these pressures would be increased where there was possibility of rejection for nonconformity.

Ss were asked to read a delinquency case study and to indicate privately an opinion position on a scale of personal vs. environmental responsibility. A false consensus gave most Ss the impression that they were deviates from a clearly defined group norm. The opinions were to remain private, though later the group would discuss the case and reach agreement on a report of the case. A later statement of opinion confirmed the hypothesis that Ss would alter their opinion to conform to the group norm even though that opinion was to remain private.

Ss were then asked to write individual descriptions of the case study. In some groups, these descriptions were to be passed around to all other Ss. In other groups the descriptions

were to remain private. Also, some groups were told that they might reject some members, while other groups were not offered this possibility.

It was hypothesized that deviates would be more likely to change their opinions toward the norm when their descriptions were to be seen by others; deviates whose descriptions were to remain private would be less likely to change opinions. The hypothesis was confirmed, even though the opinions themselves were to remain private under all conditions. It was also predicted that Ss would be more likely to change their opinions when there was possibility of rejection. This hypothesis was confirmed with qualifications. Possibility of rejection, in and of itself, did not increase change. However, Ss who did not fear rejection and whose descriptions were not to be made public conformed significantly less than the others.

It was further predicted that group pressures would operate on the content of the descriptions: the tendency to distort content so as to support the group norm would be greater for public descriptions than for private descriptions, and greater for those Ss who faced the possibility of rejection than for those who would not be rejected. These hypotheses were unconfirmed for the deviates as a whole, but differences were more clear for the Ss who held less extreme positions and were therefore less involved in their own prior opinions. For these less extreme Ss, the public descriptions were indeed significantly more supportive of the group norm than the private descriptions. The effects of possibility of rejection on distortion of content paralleled those upon change in opinion. The greatest distortion occurs when there is both possibility of rejection and where the statement will be made public. Least distortion occurs when there is neither possibility of rejection nor publicity.

There was also evidence that changes in opinion occurred during and shortly after the writing of the group descriptions. Such changes tended to bring opinions in line with the content of the descriptions—the group influence here operated first on the descriptions and then on the opinion itself. On the basis of this and other evidence, a theory of social influences on opinions was suggested: 1. Pressure to change may first operate on

content communicated, the individual selecting and distorting the content he communicates so as to avoid rejection from the group. 2. The deviate then begins to perceive selectively and distort items of content, so that he perceives an ever increasing body of content supporting the group norm, and fewer items supporting his own initial position. 3. Finally, the deviate sees a greater amount of content supporting the group norm than his own, and then changes his opinion toward conformity, thus bringing it in line with the phenomenological evidence.

REFERENCES

1. ASCH, S. E. The doctrine of suggestion, prestige, and imitation in social psychology. *Psychol. Rev.*, 1948, **55**, 250-277.
2. ASCH, S. E. Studies of independence and conformity: I. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**, No. 9 (Whole No. 416).
3. ASCH, S. E., BLOCK, H., & HERTZMAN, M. Studies in the principles of judgments and attitudes: II. Determinants of judgments by group and by ego standards. *J. soc. Psychol.*, 1940, **12**, 433-465.
4. BACK, K. The exertion of influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9-24.
5. BRUNER, J. A., & GOODMAN, C. C. Value and need as organizing factors in perceptions. *J. abnorm. soc. Psychol.*, 1947, **42**, 33-44.
6. CANTRIL, H. The intensity of an attitude. *J. abnorm. soc. Psychol.*, 1946, **41**, 129-136.
7. COCH, L., & FRENCH, J. R. P., JR. Overcoming resistance to change. *Hum. Relat.*, 1948, **1**, 512-532.
8. DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629-636.
9. DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance upon conformity to group norms. *J. abnorm. soc. Psychol.*, 1956, **53**, 100-108.
10. DUNCKER, K. Experimental modification of children's food preferences through social suggestion. *J. abnorm. soc. Psychol.*, 1938, **33**, 489-507.
11. EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
12. EVANS, J. Johnny Rocco. *J. abnorm. soc. Psychol.*, 1948, **43**, 357-383.
13. FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271-282.
14. FESTINGER, L. An analysis of compliant behavior. In M. Sherif & M. O. Wilson (Eds.), *Group relations at the crossroads*. New York: Harper, 1953. Pp. 232-255.
15. FESTINGER, L. A theory of social comparison processes. *Hum. Relat.*, 1954, **7**, 117-140.
16. FESTINGER, L., GERARD, H. B., HYMOVITCH, B., KELLEY, H. H., & RAVEN, B. The influence process in the presence of extreme deviates. *Hum. Relat.*, 1952, **4**, 327-346.
17. FESTINGER, L., & THIBAUT, J. Interpersonal communication in small groups. *J. abnorm. soc. Psychol.*, 1951, **46**, 92-100.
18. FRENCH, J. R. P., JR., & RAVEN, B. H. The bases of social power. In D. Cartwright, et al., *Studies in social power*. Ann Arbor, Mich.: Institute for Social Research, in press.
19. GERARD, H. B. The effects of different dimensions of disagreement on the communication process in small groups. *Hum. Relat.*, 1953, **6**, 249-271.
20. GERARD, H. B. The anchorage of opinions in face-to-face groups. *Hum. Relat.*, 1954, **7**, 313-326.
21. GUTTMAN, L., & SUCHMAN, E. R. Intensity and zero point for attitude analysis. *Amer. soc. Rev.*, 1947, **12**, 57-67.
22. JACKSON, J. M., & SALTZSTEIN, H. D. Group membership and conformity processes. Research Center for Group Dynamics, Univer. of Michigan 1956.
23. JANIS, I. L., & FADNER, R. The coefficient of imbalance. In H. D. Lasswell, N. Leites, et al., *Language of politics*. New York: George W. Stewart, 1949.
24. KELLEY, H. H., & SHAPIRO, M. M. An experiment on conformity to group norms where conformity is detrimental to group achievement. *Amer. sociol. Rev.*, 1954, **19**, 667-677.
25. KELLEY, H. H., & VOLKART, E. H. The resistance to change of group-anchored attitudes. *Amer. sociol. Rev.*, 1952, **17**, 453-465.
26. KELMAN, H. C. Three processes of acceptance of social influence: Compliance, identification, and internalization. *Amer. Psychologist*, 1956, **11**, 361. (Abstract)
27. KRECH, D., & CRUTCHFIELD, R. S. *Theory and problems in social psychology*. New York: McGraw-Hill, 1948.
28. LEVINE, R., CHEIN, I., & MURPHY, G. The relation of intensity of a need to the amount of perceptual distortion. *J. Psychol.*, 1942, **13**, 283-293.
29. RAVEN, B. H. The effects of group pressures on opinions, perception, and communication. Unpublished doctoral dissertation, Univer. of Michigan, 1953.
30. RAVEN, B. H., & FRENCH, J. R. P., JR. Legitimate power, coercive power, and observability in social influence. *Sociometry*, 1958, **21**, 83-97.
31. RAVEN, B. H., & RIETSEMA, J. The effects of varied clarity of group goal and group path upon the individual and his relation to his group. *Hum. Relat.*, 1957, **10**, 29-47.
32. SCHACTER, S. Deviation, rejection, and communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 190-208.
33. SCHANCK, R. L. A study of a community and its groups and institutions conceived of as behavior of individuals. *Psychol. Monogr.*, 1932, **43**, No. 2.
34. SHERIF, M. A study of some social factors in perception. *Arch. Psychol.*, 1935, No. 187.

Received October 14, 1957.

CRITIQUE AND NOTES

THE TEST ANXIETY QUESTIONNAIRE: SCORING NORMS FOR A NON-COLLEGE POPULATION¹

ZANWIL SPERBER

University of California, Los Angeles

SARASON, Gordon, and Mandler describe a Test Anxiety Questionnaire (TAQ) designed to measure anxiety specific to testing situations (2, 3, 4). The format (4, p. 810) and scoring method (3) of the TAQ require computation of the median response of a group of Ss to each of 35 questions before a given S's responses can be scored. Most of the time and effort in scoring the TAQ is expended in computing the medians. A short-cut scoring method would encourage more extensive use of a measure which has demonstrated its value as a research tool.

Sarason and Gordon present data showing the stability of medians obtained from two separate samples of Yale students (3) and suggest that these published medians might serve as norms for scoring directly from an S's protocol. However, they also caution that until it is demonstrated that other groups have medians similar to those of Yale students, the TAQ should be scored the long way with "local" norms.

In the present paper, TAQ data obtained from a noncollege sample of U. S. Air Force recruits are compared with those from Yale students. One can then begin to evaluate the general usefulness of Sarason and Gordon's suggested short-cut scoring procedure.

YALE AND AIR FORCE SUBJECTS COMPARED

The TAQ was administered to a sample of 294 male recruits in December, 1953, as part of a larger study of the effects of anxiety and stress on performance (5). In contrast to the Yale college students, a majority of the recruits did not complete high school. They had, on the average, 10.8 years of schooling ($\sigma = 1.5$). As an estimate of intelligence, each recruit's stanine score on the Technician's Specialty Index (1) was available. Although the Ss had been selected to have scores of at least 3 on this measure, which yields scores through 9, the mean was only 5.4 ($\sigma = 1.7$). The mean age of the recruits was 18.8 years ($\sigma = 3.3$), probably not very different from the average age of Yale sophomores and juniors. Approximately 15% of the recruits were Negroes.

¹ The use of USAF recruits as Ss was secured through the cooperation of Abraham Carp of the Air Force Personnel and Training Research Center. The study, however, was not sponsored by the USAF, and the Air Force is not responsible for this report.

It is also possible to compare the social-class background of the present sample and the Yale groups. The recruits filled out a face sheet which included questions about their father's education and occupation. A total of 252 recruits could supply information about their fathers' education. In contrast to the fathers of Yale Ss, 70% of whom had completed college, only 19 of the recruits' fathers had schooling beyond high school, and only four were college graduates. On the average, the fathers of the 252 recruits had 8.20 years of education ($\sigma = 3.79$).

The occupations of Yale fathers were categorized in three strata (4). Some 23% were in upper-class occupations (executives, brokers, directors of corporations, etc.); 70% were in upper-middle or middle-class occupations (managerial, professional men, owners of small firms, civil service employees, etc.); and only 7% were in the lowest class of occupations (manual laborers, secretarial, and minor civil service occupations). In contrast to the Yale fathers, the occupations of 251 recruits' fathers were distributed as follows: 110 in skilled manual labor, 56 in unskilled manual labor, 5 firemen, 22 farmers, 9 in secretarial occupations, 7 salesmen, 20 in managerial occupations (including foremen in manual labor situations), 7 in their own small business, 6 in professional occupations or owners of property or larger businesses, and 9 in miscellaneous occupations (artist, ballet teacher, politician, a "regular" in military service, etc.). At least 85% of these occupations fall in the lowest class noted by Sarason and Mandler, with, at most, 2% in the upper-class stratum and the remaining 13% in the middle-class category.

The recruit sample, for which TAQ norms will be reported, obviously differed significantly from the Yale samples with respect to social class and educational background and very probably in level of intelligence.

TAQ DATA

Table 1 presents the median response of the recruits and the two Yale samples to each of the 35 scored questions of the TAQ.² In general, the medians were similar. For 11 questions, the medians

² The recruit data was originally scored in units of 1 cm (5, p. 218), but medians were converted to the 1.5 cm unit utilized by the Yale researchers to make the results comparable.

TABLE 1

COMPARISON OF THREE SAMPLES: POINTS (IN RAW SCORE UNITS) BETWEEN WHICH THE MEDIAN RESPONSE OCCURS

Question ^b	Samples			Question ^b	Samples		
	Re-cruits, 1953 (N = 294)	Yale, 1951 ^a (N = 392)	Yale, 1952 ^a (N = 359)		Re-cruits, 1953 (N = 294)	Yale, 1951 ^a (N = 392)	Yale, 1952 ^a (N = 359)
4	4-5	3-4	3-4	23	4-5	4-5	4-5
5	5-6	3-4	3-4	24	4-5	3-4	2-3
6	4-5	3-4	3-4	25	4-5	4-5	4-5
7	4-5	4-5	4-5	26	3-4	3-4	3-4
8	3-4	2-3	2-3	27	2-3	1-2	1-2
9	3-4	4-5	4-5	28	5-6	2-3	3-4
10	4-5	2-3	2-3	29	5-6	6-7	6-7
11	4-5	3-4	3-4	30	4-5	4-5	4-5
12	1-2	2-3	2-3	31	4-5	5-6	4-5
13	0-1	0-1	0-1	32	5-6	4-5	4-5
14	2-3	2-3	2-3	33	3-4	4-5	4-5
15	4-5	3-4	3-4	34	3-4	4-5	4-5
17	5-6	4-5	4-5	35	4-5	4-5	4-5
18	4-5	4-5	4-5	36	4-5	6-7	6-7
19	3-4	4-5	4-5	37	3-4	3-4	3-4
20	5-6	4-5	4-5	38	1-2	2-3	1-2
21	5-6	4-5	4-5	39	4-5	3-4	2-3
22	4-5	4-5	4-5				

^a Reproduced from Sarason and Gordon, Table 1 (3, p. 448).

^b The scored questions are numbered as they appear on the TAQ.

TABLE 2

COMPARISON OF THREE SAMPLES: PERCENTAGE OF SUBJECTS RECEIVING VARIOUS TAQ SCORES

Yale Sample, 1951 ^a (N = 392)		Yale, Sample 1952 ^a (N = 359)		Recruit Sample, 1953 (N = 294)	
Scores included	%	Scores included	%	Scores included	%
23-35	30.8	24-35	27.6	24-35 ^b	24.8
0-12	28.6	0-12	28.1	0-11	25.9

^a Reproduced from Sarason and Gordon's Table 2 (3, p. 448).

^b The highest score possible is 35; the highest score any recruit received was 32.

of all three samples fell in the same place. For 18 additional questions, there was a deviation of but one unit, the recruit sample differing from both Yale samples by this amount on 16 of these 18 questions. For only one of the 6 remaining questions, No. 28, was there a deviation of more than two units. It would appear that when using the TAQ with male Ss, little error will be introduced if 29 of the 35 questions are scored directly from the published norms. Calculation of local norms for the remaining 6 questions (Nos. 5, 10, 24, 28, 36, and 39) would appear to be advisable.

With respect to the over-all scores the Ss received on the TAQ, the distributions of total scores are almost identical for all three samples. The median

TABLE 3

CORRELATIONS BETWEEN SCORES ON THE TAQ, WINNE, AND TAYLOR SCALES

Low Stress Ss (N = 148)			High Stress Ss (N = 146)		
Test	TAQ	Winne	Test	TAQ	Winne
Winne	+.12		Winne	+.22**	
Taylor	+.20*	+.62**	Taylor	+.35**	+.64**

* Statistically significant; $p < .05$.

** Statistically significant; $p < .01$.

score reported for both Yale samples falls between 18 and 19 (3, p. 447), and it falls at the same point for the recruit sample. Table 2 presents additional data on the distribution of total scores for the three samples. It is evident that similar cutoff scores would discriminate about the same percentage of Ss in each sample (also see 4, p. 811, Fig. 1).

The relationship between scores on a measure of anxiety specific to testing situations and measures of general anxiety level is also of interest. Gordon and Sarason (2) report a statistically significant correlation of +.468 between TAQ scores and scores on a measure of general anxiety. The recruit sample, in addition to the TAQ, also took the MMPI, which contains two measures of manifest anxiety. Ss' scores on the 50-item Taylor Manifest Anxiety Scale (6) and on the 22 items of Winne's Scale of Neuroticism (7)³ which do not overlap with Taylor's scale were obtained. These measures were developed independently, using different methods of scale construction.

Prior to taking either the TAQ or MMPI, the recruits had been tested on a battery of cognitive tasks, half under low stress and half under high stress. The product-moment correlations between scores on the three anxiety scales were calculated separately for the low and high stress groups. Table 3 presents both sets of correlations. Although most of the correlations reach levels of statistical significance with the large number of Ss involved, the correlations between the TAQ and general anxiety measures are low. The TAQ can therefore be considered as a measure of a specific anxiety. An S who is prone to become highly anxious in testing situations may not necessarily be particularly vulnerable to other kinds of stressful situations unrelated to testing.

SUMMARY

The TAQ responses of a sample of U. S. Air Force recruits, differing from Yale students with respect

³ Winne's interpretation of his results indicates that his scale can be considered a measure of manifest anxiety (7, p. 120-121).

to social class, amount of education, and intelligence level proved to be very similar to the responses of two samples of Yale college students. Median responses to 29 of the 35 questions were within one scoring unit for all three samples, indicating that a more efficient direct scoring method for these items is feasible. For all three samples, the distributions of total scores on the TAQ were almost identical. While positively related to scores on measures of general anxiety, TAQ scores appear to reflect particular sensitivity to the specific type of stress associated with testing.

REFERENCES

1. DAILEY, J. T., LECZNAR, W. B., & BROKAW, L. D. *Development of the Airman Classification Test Battery*. San Antonio, Texas: Human Resources Research Center, Lackland Air Force Base, November 1948 (Res. Bull. 48-5).
2. GORDON, E. M., & SARASON, S. B. The relationship between "test anxiety" and "other anxieties." *J. Pers.*, 1955, **23**, 317-323.
3. SARASON, S. B., & GORDON, E. M. The Test Anxiety Questionnaire: Scoring norms. *J. abnorm. soc. Psychol.*, 1953, **48**, 447-448.
4. SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 810-817.
5. SPERBER, Z. A study of the role of anxiety level and defense preference in performance under stress. Unpublished doctoral dissertation, Univer. of Michigan, 1956.
6. TAYLOR, JANET A. A Personality Scale of Manifest Anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.
7. WINNE, J. F. A scale of neuroticism. *J. clin. Psychol.*, 1951, **7**, 117-122.

Received September 30, 1957.

AUTHORITARIANISM, VERBAL ABILITY, AND RESPONSE SET¹

DONALD R. BROWN AND LOIS-ELLIN DATTA²

Bryn Mawr College

EVIDENCE that the F scale (1) was measuring variables other than authoritarianism, e.g., age, socioeconomic status, intelligence, and the form in which the items were presented, has been reported by Hyman and Sheatsley (5). The problem of item form as a source of artifacts has been investigated by a number of authors whose work has been reviewed in detail by Christie et al. (3), who also report data gathered with a carefully devised scale of reversed F items. The present study was designed to clarify the meaning of observed changes in F scores during four years of a liberal arts college experience as these changes relate to possible artifacts arising from such factors as verbal ability, educational level, and response set.

Existing evidence clearly suggests that some of the variance on the original F scale can be attributed to such artifacts, particularly in groups that are not extreme on F. However, it was our thought that a more subtle instrument, derived empirically with the F scale as a criterion, might

not be so open to this type of criticism. An instrument of this kind, the F₄ scale, has been developed by Webster et al. (8, 9, 11). In addition, a scale for measuring response set (9), and a method for removing the variance of such a response set scale (10), have been reported by Webster and are described in the next sections. We have also hypothesized, in line with the findings of Brown and Bystryn (2) and of Sanford (6), that education effects real changes in personal-cognitive modes of adaptation, and that these are reflected in F₄ scores and are not mere artifacts of verbal ability or sophistication in a purely genteel sense.

METHOD

Variables

In our analysis we explored the interactions of authoritarianism as measured by the F₄ scale, verbal ability (SAT-V) as measured by the verbal aptitude score on the scholastic aptitude section of the College Entrance Examination Boards taken during the last year of high school, the educational level of the Ss as measured by year in college, and response set as measured by the S's score on a scale developed, according to the method of Fricke (4, 10), from a pool of 677 items of which F₄ is a part.

¹ The data reported here were collected at Vassar College as part of the research program of the Mary Conover Mellon Foundation. The authors wish to express their appreciation to Nevitt Sanford, Harold Webster, Robert S. Davidson, and Mary Mehl for their assistance in gathering and interpreting the data.

² Now with Courtney & Co., Philadelphia, Pa.

TABLE 1

UPPER AND LOWER LIMITS FOR THE UPPER, MIDDLE, AND LOWER THIRDS OF THE SAT-V AND R₂ SCALE DISTRIBUTIONS FOR FRESHMEN AND SENIORS IN THE SAMPLE OF 512 SUBJECTS

Group	R ₂			SAT-V		
	Lower	Middle	Upper	Lower	Middle	Upper
Freshmen	6-33	34-49	50-84	430-544	545-578	579-720
Seniors	6-34	35-48	49-87	390-537	538-595	596-750

TABLE 2

F₄ SCORE MEANS FOR 135 COLLEGE FRESHMEN AND 135 SENIORS DIVIDED INTO THIRDS ON THE BASIS OF THEIR R₂ AND SAT-V SCORES

R ₂	SAT-V			
	Upper	Middle	Lower	Mean
Upper	Fr. 64.5 Sr. 56.4	68.7 60.5	86.7 68.9	73.3 61.9
Middle	Fr. 55.7 Sr. 50.8	61.7 51.9	68.3 61.3	61.9 54.7
Lower	Fr. 47.7 Sr. 35.9	54.2 45.9	58.5 48.1	53.5 43.3
Mean	Fr. 56.0 Sr. 47.7	61.5 52.8	71.2 59.4	

Briefly, this scale of response set was based on the assumption that true-false items which are responded to in a 50/50 split are more likely than other items to have an ambiguous content. Ss who score at the extremes on a scale constructed of such items should therefore be distinguished by response in terms of some internal tendency to answer positively to ambiguous statements. The present response set scale consists of 105 items all scored "true" with means ranging from .32-.54 (Type II Error B < .01). Its KR 21 reliability was .91 for 512 Ss.

Subjects

From our original sample of over 1500 freshmen and seniors tested at a women's liberal arts college, a random sample of 512 students (256 freshmen and 256 seniors) was drawn for experimental study by the staff of the Mellon Foundation. From the scores of this sample on SAT-V and R₂, four grouped frequency distributions were made in order to choose, for a comparison in terms of F₄ scores, Ss who fell in the upper, middle, and lower thirds of the SAT-V and R₂ distributions. Table 1 describes these distributions for the sample of 512. From this stratified sample, 15 cases were randomly selected from

TABLE 3

CORRECTED F₄ SCORE MEANS FOR 135 COLLEGE FRESHMEN AND 135 SENIORS DIVIDED INTO THIRDS ON THE BASIS OF THEIR R₂ AND SAT-V SCORES

R ₂	SAT-V			
	Upper	Middle	Lower	Mean
Upper	Fr. 33.3 Sr. 25.7	36.8 29.7	51.3 37.3	40.5 30.9
Middle	Fr. 33.6 Sr. 29.3	39.3 30.7	46.5 38.1	39.8 32.7
Lower	Fr. 34.9 Sr. 22.7	42.7 32.8	45.9 36.3	41.2 30.6
Mean	Fr. 33.9 Sr. 25.9	39.6 31.1	47.9 37.2	40.5 31.4

each of the 18 possible combinations of variables we wished to study by an analysis of variance design (for example, 15 Ss who were seniors in the upper third of the senior SAT-V distribution and the upper third of the R₂ distribution etc.) The groups were thus equalized in terms of size in order to maximize the reliability of the estimated interaction analysis.

RESULTS

The 270 F₄ scores were then corrected for response set by the method proposed by Webster (10). Briefly, this method consists of obtaining a score *Y* uncorrelated with a suppressor variable *t* by weighting the variable *t* by the regression coefficient *k* of the variable *t* on *T* and subtracting the result from the original score *T*. Thus each score on F₄ can be corrected for R₂ by the formula $Y = F_4 - (k)R_2$ where *k* = the linear regression of R₂ on F₄.

Tables 2 and 3 give the resulting distribution of F₄ and corrected F₄ scores with means for freshmen and seniors in the sample. It can be seen that seniors score lower than freshmen on F₄ and on corrected F₄; that F₄ varies positively with R₂ and inversely with SAT-V.

The data of Table 2 were subjected to two analyses of variance; one for F₄ and one for corrected F₄ scores. Table 4 gives the results for F₄ scores and Table 5 those for corrected F₄ scores. In Table 4 for the F₄ scores the *F* ratio of the R₂ and the within groups mean square³ was 48.36; for the SAT-V, 35.22; for year in college, 24.18.

³ Since Bartlett's test for homogeneity of variance when applied to the F₄ categories for the sample of 270 yielded a χ^2 of 6.4531 which is not significant, it seemed safe to attribute significant *F* ratios to differences in subgroups means.

TABLE 4

ANALYSIS OF VARIANCES FOR A SAMPLE OF 270 STUDENTS WITH YEAR IN COLLEGE, SAT-V, AND R_2 SCORES AS THE VARIABLES WHOSE EFFECTS ON F_4 ARE TO BE EVALUATED

Source	Sum Squares	df	Mean Square	F
Response set	16635	2	8318	48.36 ^a
Year in college	6059	1	6059	35.22 ^b
SAT-V	8318	2	4159	24.18 ^a
$R_2 \times$ SAT-V	814	4	204	1.18 ^c
$R_2 \times$ Year	194	2	97	—
SAT-V \times Year	36	2	18	—
Year \times SAT-V \times R_2	386	4	97	—
(Between groups)	(32442)	(17)		
Within groups	43292	252	172	
Total	75734	269		

^a $F_{.001} = 7.31$ with df gms 2 and sms 120.

^b $F_{.001} = 11.38$ with df gms 1 and sms 120.

^c $F_{.05} = 2.45$ with df gms 4 and sms 120.

TABLE 5

ANALYSIS OF VARIANCE FOR A SAMPLE OF 270 STUDENTS WITH YEAR IN COLLEGE, SAT-V, AND R_2 SCORES AS THE VARIABLES WHOSE EFFECTS ON CORRECTED F_4 SCORES ARE TO BE EVALUATED

	Sum Squares	df	Mean Squares	F
Response set	16	2	8	22.29 ^a
SAT-V	7241	2	3620.5	
Year in college	5549	1	5549	24.17 ^b
$R_2 \times$ SAT-V	570	4	142.5	
$R_2 \times$ Year	141	2	70.5	
Year \times SAT-V	88	2	44	
Year \times SAT-V \times R_2	252	4	63	
(Between groups)	(13857)	(17)		
Within groups	40922	252	162.4	
Total	54779	269		

^a $F_{.001} = 7.31$ with df gms 2 and sms 120.

^b $F_{.001} = 11.38$ with df gms 1 and sms 120.

The above are all significant at better than the .001 level. No other interactions were significant.

Inspection of Table 5 shows that the relation of R_2 to corrected F_4 disappears as is to be expected, since it is treated as a suppressor. Verbal ability (SAT-V) and year in college remain significantly related to corrected F_4 at the .001 level. All other interactions fail to achieve significant F ratios.

DISCUSSION

The fact that verbal ability and year in college, both high correlates of intelligence, remain significantly related to F_4 , even with response set

removed, but are not related to response set in the same degree, supports the contention that both of these variables and authoritarianism are related in a more basic way than as a simple reflection of acquiescence. Christie's (3) work with reversed items of the original F scale further supports this contention.

It would appear, then, that neither verbal ability nor year in college accounts for high authoritarianism scores through the lack of discernment on the part of the S in the face of verbal material. Indeed, the striking fact of a lack of significant relationship between response set and year in college in our findings indicates that response set measures a basic disposition to structure ambiguity in a creative process which permits agreement rather than rejection by the S . Perhaps this is another reflection of the general superficial tolerance in the face of controversy so often attributed to the present generation of students. The decrease of F_4 with year in college can then be explained in line with the similar findings on development in college or a real increase in liberal and intellectual tendencies independently of simple verbal facilitation, sophistication or poise, and underlying disposition to be agreeable. In less well educated groups, on the other hand, response set may merely be a way of defending against lack of comprehension.

SUMMARY

Two hundred and seventy college women, half freshmen and half seniors, were administered the F_4 and R_2 scales as a part of a larger test battery. The scores were analyzed by an analysis of variance design to explore the relationships among F_4 scores corrected for response set, year in college, and verbal ability. The criticism of the authoritarian syndrome on the basis of acquiescence is challenged by the results.

REFERENCES

1. ADORNO, T. W., FRENKEL-BRUNSWICK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
2. BROWN, D. R., & BYSTRYN, B. College environment, personality, and social ideology of three ethnic groups. *J. soc. Psychol.*, 1956, **44**, 279-287.
3. CHRISTIE, R., HAVEL, J., & SEDENBERG, B. Is the F scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143-159.
4. FRICKE, B. G. Response set vs. a suppressor variable in the OASIS and MMPI. *J. consult. Psychol.*, 1956, **20**, 161-169.
5. HYMAN, H. H., & SHEATSLEY, R. S. The authoritarian personality—a methodological critique. In R. Christie & Marie Jahoda (Eds.), *Studies in the scope and method of the authoritarian personality*. Glencoe, Ill.: Free Press, 1954.

6. SANFORD, N. (Ed.) Personality development during the college years. *J. soc. Issues*: 1956, **12**, No. 4.
7. STAGNER, R. A review of psychology of personality: Six modern approaches by J. J. McCary. *Contemp. Psychol.*, 1957, **2**, 42-43.
8. WEBSTER, H. Some quantitative results. In H. Sanford (Ed.), Personality development during the college years. *J. soc. Issues*, 1956, **12**, 29-43.
9. WEBSTER, H. *Research manual for VC Attitude Inventory and VC Figure Preference Test*. Vassar College: Mary Conover Mellon Foundation, 1957.
10. WEBSTER, H. Correcting personality scales for response sets or suppression effects. *Psychol. Bull.*, 1958, **55**, 62-64.
11. WEBSTER, H. SANFORD, N., & FREEDMAN, M. A new instrument for measuring authoritarianism in personality. *J. Psychol.*, 1955, **40**, 73-84.

Received September 25, 1957.

INDIVIDUAL VERSUS GROUP GOAL CONFLICT^{1, 2, 3}

EWART E. SMITH

Fels Group Dynamics Center, University of Delaware

BOTH casual observation and research suggest that an important determiner of group efficiency and adaptiveness is how group members resolve conflicts between individual and group goals. Deutsch (3) has demonstrated the importance of cooperative behavior in effective groups. Mintz (4), using an ingenious group task that required cooperation for success, found that increasing individual motivation frequently decreased cooperation and therefore lowered efficiency. He theorizes that uncooperative behavior, being nonadaptive, does not occur in a group unless the cooperative pattern is broken by the uncooperative deviate. Once the cooperative pattern is disturbed, however, cooperation is no longer rewarding, and nonadaptive competition rapidly develops.

Mintz appears to postulate that the forces arising from the norms of internalized reference groups are sufficiently strong, in most individuals, to counteract individual, competitive motivations. The contention here, however, is that such internalized

group forces are not strong enough to counteract the strong individual goal forces that are aroused in many real situations, such as Mintz's example of the theatre fire; strong individual goal behavior can only be successfully opposed by immediate visible external group forces.

These considerations lead to the following hypotheses:

1. When faced with an alternative, persons will choose the attainment of their own goal in preference to that of a group goal less frequently in an overt situation, in which their choice is known by the group, than in a covert situation where their choice is secret.

2. When faced with an alternative, persons will choose the attainment of their own goal in preference to that of a group goal more frequently under increased individual motivation than under reduced individual motivation. The first hypothesis is consistent with the powerful effects of groups demonstrated in research on conformity, such as that by Asch (1). It is inconsistent, however, with Mintz's position, which implies that in a covert situation most individuals should be cooperative, as the infrequent uncooperative act would not be perceived by the group and the cooperative pattern should not, therefore, be disturbed. The second hypothesis appears to be self evident and is consistent with Mintz's data.

METHOD

Subjects

The subjects (Ss) were 120 male flying personnel, including 92 officers and 28 enlisted men, who were attending the Survival School at Stead Air Force Base. The six-man experimental groups were actual Air Force crews. Some crews had been together for many months; others consisted of men assigned to the same crew for their stay at the

¹ This report is based on work done under ARDC Project No. 7723, Task No. 77461, in support of the research and development program of the Air Force Personnel and Training Research Center, Lackland Air Force Base, Texas. Permission is granted for reproduction, translation, publication, use and disposal in whole or in part by or for the United States Government. The opinions or conclusions expressed or implied herein are those of the author. They are not to be construed as necessarily reflecting the views or endorsement of the Department of the Air Force or of the Air Research and Development Command.

² The writer is indebted to Carl W. Backman, Phil W. Buck, Willard F. Day, Robert McQueen, Edwin H. Richardson, Paul F. Secord, Walter A. S. Smith, E. Paul Torrance, and to Charles E. Hawkins, who served as experimenters and gave helpful criticisms and advice.

³ This study was conducted while the author was with the Survival Methods Branch, Air Force Personnel and Training Research Center, Stead Air Force Base.

Survival School. All crews had been living together for a minimum of five days.

Procedure

Two independent variables were used; the first was high individualistic motivation, induced by an evaluation set, versus low individualistic motivation. The second independent variable involved comparison of an overt situation, in which the group members knew when an *S* failed to sacrifice his solution to help the group, with a covert situation in which the group members did not know when an *S* failed to sacrifice. There were thus four conditions (with five six-man crews in each): (a) high individualistic motivation, overt; (b) high individualistic motivation, covert; (c) low individualistic motivation, overt; and (d) low individualistic motivation, covert. Ten experimenters were used.

The procedures, instructions, and apparatus⁴ were copied from Crutchfield's assessment technique, with minor modifications. The reader is referred to Crutchfield (2) for a detailed description. The *Ss* were seated in a circle with their backs to one another. Each *S* was to assemble a square with some geometric pieces. As the initial pieces held by each *S* did not form a square, the *Ss* had to request and exchange pieces by means of a tray carried around the group by the experimenter. The instructions indicated that it was easy for one or two people to form a square, but difficult for everyone to have a square simultaneously. The *Ss* were told that at the end of the allotted time (unspecified), if they each have a square, the group would receive 30 points (i.e., 5 points each). If they did not each have a square, those who did would each receive 5 points.

By manipulating the tray, the experimenter was able to control the situation. Each *S* passed through four identical rounds. In the first round he requested a needed piece; in the second, he received it and completed his square; in the third and fourth rounds he held his square unmodified since there was no request for a piece held by him. The stage was now set for the critical trials, where stress was placed upon each *S* by facing him with a request for a piece which was part of his completed square. These critical trials were continued until each *S* had given up the requested piece, or until there had been sixteen critical trials.

The basic instructions and procedures were varied in the four experimental conditions, in the following manner:

A. High individualistic motivation conditions (both overt and covert): The instructions stated

⁴ The only modification in the apparatus was the use of the word "request" on the underside of the request pieces.

that the experimenters had been asked to evaluate the intellectual aptitude of the *Ss* and would do so by means of a test. They were told that the results would be sent to their commanding officers, who would discuss the results with them.

B. Low individualistic motivation conditions (both overt and covert): The instructions stated that the task was an experimental one, not known to measure anything. The *Ss* were told that it would not be necessary to give their names.

C. Covert conditions (high individualistic motivation and low individualistic motivation): The following sentences were added to the basic instructions:

We have provided you each with a box to make sure that no one sees the solution to the problem by looking at all the pieces at once, or even just at those of a person to his side. However, since the boxes don't cover up the pieces entirely, you should look straight ahead and not to either side.

On any given presentation you may either simply pass, that is, by shaking your head indicate to the messenger that you neither want to exchange nor to request a piece, or both exchange and request.

D. Overt conditions (high individualistic motivation and low individualistic motivation): In these conditions, the basic instructions were modified as follows:

On any given presentation you may either simply say "Pass" indicating to the messenger that you neither want to exchange nor to request a piece, or you may exchange one piece or request one piece, or both exchange and request.

In addition, when the experimenter presented the tray on a critical trial he said, "Someone needs one of your pieces." If the *S* passed, the experimenter repeated after him, "You pass."

An *S*'s score was the trial on which he gave up the requested piece.

The Design in Summary

The procedure described was, in summary, intended to create the following psychological conditions:

High individualistic motivation. The individual was in danger of losing status, in his own mind and that of his commanding officer, because of failure on an intellectual task. At the same time, he was confronted with requests from others for help. To help the others would increase the possibility that the individual himself would fail.

Low individualistic motivation. There was little extrinsic motivation for attempting to do well. The results would be anonymous, and the task was not known to measure anything. However, we may assume some minimal motivation to do well. Again

TABLE 1
NUMBER OF SUBJECTS WHO SACRIFICED ON EACH
CRITICAL TRIAL IN FOUR EXPERIMENTAL
CONDITIONS

Critical Trials	High Individualistic Motivation Overt	Low Individualistic Motivation Overt	High Individualistic Motivation Covert	Low Individualistic Motivation Covert
1-4	20	25	2	4
5-8	2	1	2	6
9-12	0	1	2	4
13-16	0	0	2	4
Never Sacrificed	8	3	22	12

Note.— $N = 30$ in each condition.

TABLE 2
NUMBER OF SUBJECTS WHO SACRIFICED IN FOUR
EXPERIMENTAL CONDITIONS

	Overt	Covert	Totals
High individualistic motivation	22	8	30
Low individualistic motivation	27	18	45
Totals	49	26	

Note.— $N = 30$ in each condition.

the individuals were in conflict between their motivations to do well and their desire to help their associates.

Overt conditions. The individual knew that his decision, as to how he resolved his conflict, would be known by the group.

Covert conditions. The individual knew that the group would not know how he resolved his conflict. The covert situation was only relatively covert, however, as the experimenter obviously knew how the S resolved it. The S s appeared to feel under pressure from the experimenter to help their group as indicated in the S s who had not sacrificed by such behavior as ignoring the presence of the experimenter and the tray on subsequent critical trials, avoidance of eye contact with the experimenter, etc.

Observer reports and interviews with S s indicated that the experimental manipulations produced the desired psychological conditions.

RESULTS

The data in Table 1 indicate when the S s, in each condition, broke their squares, thus jeopardizing their individual solution to help their group and, in effect, temporarily giving up their own goals in favor of group goals. As the distributions are skewed, the data have been simplified to permit a chi square analysis (see Table 2.)

The chi square on the number of S s in all the

covert groups who broke their squares, compared to the number in all the overt groups, was 18.81, significant at the .01 level. These data support the first hypothesis.

The chi square on the number of S s breaking their squares in all high individualistic motivation groups, compared to the number in all low individualistic motivation groups, was 8.00, significant at the .01 level. These results appear to support the second hypothesis. However, inspection of the data indicates a possible interaction effect between the variables of overt-covert and individualistic motivation. Interaction was tested by comparing the S s in the high individualistic motivation-covert groups with the S s in the low individualistic motivation-covert groups. The resulting chi square was 6.79, significant at the .01 level. In addition, the high individualistic motivation-overt groups were compared with the low individualistic motivation-overt groups. The resulting chi square of 1.78, was not statistically significant. Apparently, then, there was a significant interaction effect between individualistic motivation and covert-overt conditions, with the variable of individualistic motivation being potent only in the covert groups.

An attempt was made to relate cohesiveness to the readiness of S s to jeopardize their own solutions to help their fellow group members. On the assumption that the longer crew members had been together the more cohesive they would be, the number of months each S had served with the other members of his crew was correlated with readiness to break his square. No significant relationship was found.⁵

An interesting *post hoc* finding⁶ is seen in Table 1, in the striking difference between the overt and covert conditions in the number of critical trials occurring before S s broke their squares. In the overt condition, all but one of the S s who broke did so in the first half of the 16 critical trials. In the covert conditions, 14 broke in the first half of the critical trials, and 12 in the last half. A chi square comparison of the overt and covert conditions on breaking in the first eight trials versus breaking in the last eight trials is 23.07, which is highly significant. These data suggest that in the overt conditions, those S s who refused to break their squares in the first few trials had made a public decision which was difficult to change without public ad-

⁵ Subsequently, data were collected on additional crews, using as a measure of cohesiveness the degree of their desire to remain in the same crew during the arduous survival trek phase of their training. Again there was no relationship between cohesiveness and willingness to sacrifice one's own goal for the group goal.

⁶ This analysis, and interpretation, was suggested to the writer by John T. Lanzetta.

mission of wrong-doing, whereas in the covert conditions their decisions were not public and presumably could be more easily changed. If this interpretation is correct, it serves to explain a striking phenomena observed by the experimenters. Members of the covert groups evidenced high tension by signs such as perspiring and avoidance behavior, whereas members of the overt groups were relatively calm. We may suppose that in the overt groups early decisions ended the conflict, while continued freedom to change prolonged the conflict in the covert groups, with resultant higher tension levels.

DISCUSSION

The relative lack of cooperative behavior observed in the covert groups, compared to the overt groups, is contrary to Mintz's (4) postulate that uncooperative group behavior is due to the perception by the group of an uncooperative act on the part of a deviant individual. The Ss in the covert groups did not know that some of the others were being uncooperative. In the overt groups, however, at least one member of each group was openly uncooperative on the first trial, yet most Ss soon cooperated with the group.

The ineffectiveness of individualistic motivation in the overt conditions is contrary to Mintz's (4) results. Mintz found, in what was in effect an overt situation, that the addition of small monetary rewards and punishments produced an increase in uncooperative, individual-oriented behavior. This inconsistency may be due to the lack of comparability of the laboratory and field conditions.⁷

In Mintz's research, causing others to fail could, at most, result in their paying a ten-cent fine, whereas in the present experiment, failure could have far reaching effects on the other crew members' military careers. In addition, Mintz's groups were ephemeral in contrast to the real crews used

⁷ Mintz (4) states that his conclusions are tentative until verified by field data.

here who would have to work and live together after the test.

The practical implications of this research are clear. When it is desirable that persons be socially rather than individually oriented, the wise course is to structure the situation so that most behavior is open to inspection by the group.

SUMMARY

An experiment was performed in which a conflict was produced between individual and group goal attainment. An overt situation was compared to a covert situation, and a high individualistic motivation condition with a low individualistic motivation condition. The hypotheses were:

1. When faced with an alternative, persons will choose the attainment of their own goal in preference to that of a group goal less frequently in an overt situation, in which their choice is known by the group, than in a covert situation where their choice is secret.

2. When faced with an alternative, persons will choose the attainment of their own goal in preference to that of a group goal more frequently under increased individualistic motivation than under reduced individualistic motivation.

The first hypothesis was supported. The second hypothesis was found to hold only in covert situations.

A *post hoc* finding that individuals apparently feel freer to change secret decisions than public decisions is discussed.

REFERENCES

1. ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
2. CRUTCHFIELD, R. S. Assessment of persons through a quasi group interaction technique. *J. abnorm. soc. psychol.*, 1951, **46**, 577-588.
3. DEUTSCH, M. An experimental study of the effects of cooperation and competition upon group processes. *Hum. Relat.*, 1949, **2**, 199-231.
4. MINTZ, A. Non-adaptive group behavior. *J. abnorm. soc. psychol.*, 1951, **46**, 150-159.

Received September 30, 1957.

SEMANTIC ASPECTS OF PROGNOSIS¹

SPIRO B. MITSOS

Evansville State Hospital

THE measurement of meaning, though obviously important, is beset with difficulties (3). Jones and Thurstone (1) imply that these difficulties may result from failure to restrict the semantic context. They state:

¹ Part of this paper was read at the American Psychological Association meeting in 1957.

It is probably quite true that a word has no unique meaning or, more properly, that the meaning of a word depends upon the context in which it is presented. In the latter sense, a word has an infinite number of meanings each corresponding to a particular context. If such is the case, it is not possible to determine, either logically, or experimentally, the generalized meaning of a

TABLE 1
CONCORDANCE WITHIN GROUPS

Group	χ^2	p
Aide (I)	252.68	.001
Aide (II)	218.61	.001
Professional staff	134.94	.001

TABLE 2
AGREEMENT BETWEEN GROUPS

Groups	r	p
Aide (I) vs. Aide (II)	.90	.001
Professional vs. Aide (I)	.67	.001
Professional vs. Aide (II)	.79	.001

word. However, it may be possible to present words in a particular context and to determine their meaning in terms of that imposed context (1, p. 31).

In their study, Jones and Thurstone presented a list of descriptive adjectives on a successive interval schedule. Subjects were asked to indicate, along a nine-point scale, the meaning of each word or phrase in terms of the degree to which each denoted like or dislike for food.

The present study stems from the work of Jones and Thurstone in that a psychophysical scaling procedure, the rank-order technique, was applied to a problem of word meanings in a restricted semantic context, that of the communication of behavioral descriptions between and within professional and nonprofessional groups in a general psychiatric hospital. The essential commonality and specific differences in meaning of psychiatric symptom terms were assessed among psychiatric aides and professional staff members in the specific semantic context of prognostication. Predictions were as follows:

1. Within the semantic context of prognostication, psychiatric symptom terms evoke significant commonality of meaning (a) among professional staff members (psychiatrists and psychologists), (b) among psychiatric aides, and (c) between professional staff members and psychiatric aides.

2. Prolonged personal contact between professional staff members and psychiatric aides increases the commonality of meaning of terminology.

3. In the case of terms on which aides and professionals tend to disagree, the aides perceive those terms that have high personal threat value for them as contributing toward a poorer prognosis than do the professional staff members.

METHOD

Subjects (Ss) for this experiment included 12 professional staff members and 35 psychiatric aides from a small state mental hospital. The professional staff were, for the most part, recent additions to the staff of the hospital, which was then undergoing extensive reorientation toward education and active treatment programs. Consulting relationships between professional staff and groups of ward personnel had recently been established, thus greatly increasing contact and communication between professionals and aides.

A list was compiled of 30 psychiatric symptom terms and phrases commonly used by both groups of Ss. Each item appears on at least one of the ward behavioral charts regularly used by aides. Ss were asked to rate independently each of the 30 items on a 10-point scale ranging from "least serious" to "most serious" in terms of the degree to which each contributes toward a favorable prognosis. The distribution of rankings was forced, each S being compelled to rank three items at each of the 10 scale points, thus yielding a modified rank-order distribution. After a six-month interval, the aides repeated this procedure.

RESULTS AND DISCUSSIONS

Two professional staff members and eight aides had to be dropped from the study because of failure to follow instructions on the forced ranking. The data reported, therefore, are based on the remaining 37 Ss.

Within each set of ratings, frequency tables were constructed by tabulating the number of times each symptom term was ranked at each scale point. In order to treat the data by rank-order technique, each of the 10 scale points was regarded as a three-way tie in a distribution of 30 items. Chi squares of concordance were then computed using the formula of Friedman's "problem of m rankings" (2). These results are reported in Table 1. Each value is significant at well beyond the .001 level, indicating a high degree of rater agreement within each set of ratings. These data confirm the first prediction with respect to commonality of meaning within groups.

For each set of ratings—first and second administrations to aides and single administration to professional staff—item means for the symptom terms were computed and ranked. Agreement between groups was measured by rank-order correlations, reported in Table 2. All these coefficients are significant beyond the .001 level. As predicted, the coefficient between the professional staff members and the aides on the second administration is larger than on the first, but the difference falls short of acceptable significance.

The item means and ranks are presented in Table 3. The rank position of each symptom term was compared over the three sets and the three possible rank discrepancy scores computed for

TABLE 3
MEAN RANK AND POSITION FOR EACH
GROUP OF RATERS

Item	Professionals		Aide I		Aide II	
	Mean	Rank	Mean	Rank	Mean	Rank
1. Delusions of persecution	8.80	1	7.63	3	8.41	3
2. Attempted homicide	8.50	2	9.41	1	9.18	2
^a 3. Mute	8.00	3	5.11	18	6.18	11
^a 4. Apathetic	7.90	4.5	6.04	9	5.47	14.5
^a 5. Unable to care for personal needs	7.90	4.5	5.26	17	5.47	14.5
6. Hallucinations	7.80	6	6.92	6	7.47	6
^b 7. Attempted suicide	7.60	7.5	9.18	2	9.29	1
8. Delusions of grandeur	7.60	7.5	5.78	11	6.53	9
^a 9. Incontinent	7.20	9	5.63	13	5.18	16.5
10. Seclusive	6.70	10	5.33	15.5	7.06	8
^b 11. Combative	6.00	11	7.11	4	7.24	7
12. Very prayerful	5.90	13	5.33	15.5	5.70	13
13. Destructive	5.90	13	6.89	7	6.59	10
^b 14. Sexually provocative	5.90	13	7.00	5	7.82	4
15. Dirty, Disorderly	5.70	15	4.74	21	5.18	16.5
^a 16. Very active	5.30	16	4.11	26	4.18	22.5
^a 17. Depressed	5.20	17	6.81	8	7.59	5
^a 18. Unsociable	5.00	18.5	4.22	25	5.12	18
^a 19. Quiet	5.00	18.5	3.74	27	3.18	27
20. Somatic complaints	4.70	20	4.70	22	4.35	20.5
21. Masturbation	4.10	21.5	4.96	20	4.00	26
22. Resistive	4.10	21.5	5.78	11	4.82	19
23. Excited	4.00	23	4.67	23	4.18	22.5
^a 24. Apprehensive	3.80	24	5.78	11	5.82	12
25. Aggressive	3.50	25.5	4.33	24	4.12	24
^b 26. Attempted escape	3.50	25.5	5.07	19	4.06	25
^b 27. Disobedient	3.10	27	5.41	14	4.35	20.5
28. Talkative	2.70	28	2.59	29	2.18	29
29. Profanity	1.90	29	3.67	28	2.53	28
30. Happy	1.70	30	1.78	30	1.76	30

^a Items with significant disagreement, judged low personal threat.

^b Items with significant disagreement, judged high personal threat.

each item, with a possible range of -29 to +29 and an actual range of -13 to +15. All discrepancies greater than ± 6.5 fall outside the .05 confidence interval around the median value. There were 27 such scores representing 15 symptom terms. These 15 items (indicated on Table 3) represent "significant" disagreement between groups of judges.

Each of five clinical psychologists was asked to consider these "significant" items and check those considered to possess high personal threat value for the psychiatric aide. Six of the 15 symptom terms were regarded as threatening by four of the five judges—the arbitrary criterion. The interrater tetrachoric coefficient of agreement was .78.

As predicted, the six high personal threat items were rated higher by the aides. Seven of the nine

TABLE 4
AGREEMENT BETWEEN STUDENTS AND OTHER
GROUPS

Groups	<i>r</i>	<i>p</i>
Students vs. Professionals	.60	.001
Students vs. Aide (I)	.77	.001
Students vs. Aide (II)	.76	.001

TABLE 5
NUMBER OF ACTIVE AND PASSIVE ITEMS VIEWED AS
MORE SERIOUS OR LESS SERIOUS BY THE
PROFESSIONAL STAFF

	Active	Passive	χ^2	<i>p</i>
Professional staff rating less serious than non-professional	12.5	2.5	96.44	.001
Professional staff rating more serious	3	12		

remaining items were rated higher by the professional staff. With 13 of the 15 items conforming to the prediction, the binomial expansion is significant beyond the .01 level. These data suggest that the meaning that persons attribute to behavioral descriptions is predictably influenced by the degree to which the behavior makes them feel uncomfortable. Behaviors thought by experts to elicit threat reactions from psychiatric aides are regarded by the aides as contributing toward a poorer prognosis than would be assigned by professional staff members. There is reason to believe that, in a therapeutic milieu, a worker's behavior toward a patient will be influenced by his prognostic predictions. Such phenomena have been demonstrated for example in studies of mutual withdrawal (4).

The question may well be raised as to whether the semantic norm reflected in these data pertains to a psychiatric hospital or to the general English-speaking community. The interpretation of items that elicited disagreement in terms of personal threat is also open to question: perhaps what differentiates them from other items is simply that they reflect overt large muscle activity. A group of nonpsychiatric raters was therefore obtained consisting of 30 undergraduate students enrolled in the first course in psychology. After discussing prognostication, each *S* completed the ratings of the 30 symptom terms.

Statistical analysis of these data paralleled that previously discussed. The chi square of 301.7 indicates significant concordance within the group. Rank-order correlations computed between this set of ratings and the initial sets are all significant. (See Table 4.) In short, commonality is extended, and a general English semantic norm seems indicated.

Rank discrepancies were computed between the professionals and the students, and examined in terms of the previously established confidence interval. Without exception, where the professional staff ranked an item as more serious than the aides, they also ranked it as more serious than the students. The professionals thus appear to diverge from the general semantic norm. The aides and college students agree with each other to a greater degree than either agree with the professional staff.

To check whether the type of activity implied by the symptom term associated with the disagreements that occurred, the list of 30 items was given to four judges with instructions to designate each item as active or passive with regard to voluntary overt large muscle activity. Applying the criterion of three agreements, 15 items were designated as active and 15 as passive.

The rank positions of each item as assigned by the three nonprofessional groups were averaged and these figures compared with the rankings of the professionals. These data, reported in Table 5, indicate a sharp difference between professional and nonprofessional raters. Active items were rated consistently as more serious by the aides and students, while passive items were so rated by the psychologists and psychiatrists. The essential dichotomy reflected supports the view that the groups of raters were differentially influenced by the type of activity implied by the symptom terms.

SUMMARY

Commonality of meaning of psychiatric symptom terms among psychiatric aides and professional staff members was evaluated in the semantic context of prognostication. Groups of psychiatric aides and professional hospital personnel ranked 30 symptom terms along a 10-point scale in terms of

the degree to which each contributes toward favorable prognosis. Commonality of meaning was demonstrated between and within groups, but exceptions occurred as certain symptoms considered threatening to the aide were rated as more serious by aides.

Integration of these data with ratings completed by college students raised the hypothesis of a general English semantic norm for psychiatric symptom terms. This hypothesis was supported with all measures of commonality achieving significance. As indicated by specific differences between groups, the highly trained professional workers tend to diverge from the general semantic norm.

This divergence of the professional subjects is highly related to the active-passive dimension of the behaviors rated. Items implying overt large muscle activity were consistently viewed as contributing toward a poorer prognosis by the nonprofessional groups of raters. One would wonder to what degree these data reflect differences in "knowledge," as opposed to differences in attitudes regarding prognostication, psychopathology, and human behavior in general.

REFERENCES

1. JONES, L. V., & THURSTONE, L. L. The psychophysics of semantics: An experimental investigation. *J. appl. Psychol.*, 1955, **39**, 31-36.
2. LINDZEY, GARDNER. (Ed.) *Handbook of social psychology*. Cambridge: Addison-Wesley, 1954.
3. OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univ. Illinois Press, 1957.
4. TUDOR, G. E. A sociopsychiatric nursing approach to intervention in a problem of mutual withdrawal on a mental hospital ward. *Psychiatry*, 1952, **15**, 193-217.

Received October 3, 1957.

TIME JUDGMENT, AESTHETIC PREFERENCE, AND NEED FOR ACHIEVEMENT¹

HELEN B. GREEN AND ROBERT H. KNAPP

Wesleyan University

IN THE following report, certain measures of time judgment are studied in relation to a pattern of aesthetic preferences among Scottish tartans which has been es-

tablished for McClelland's measure of Achievement (2). The present study attempts to demonstrate that performance in time judgments is related to asceticism of aesthetic taste, which has already been shown to correlate with achievement motivation. The subjects (Ss) consisted of 29 teachers ranging in age from 25 to 45 years who were enrolled in summer school at Wesleyan University.

¹ The research reported in this article was supported in part by a grant by the Behavioral Sciences Program of the Ford Foundation to David C. McClelland for study of the relationship between achievement motivation and economic development.

They proved quite heterogeneous with respect to background and educational interests. Each *S* was given three tests as follows:

1. *The Tube Test*. This test was designed to measure the *S*'s capacity to estimate the time required for a constantly moving point to reach a fixed mark after its rate of progress had been briefly observed. For this purpose, an apparatus was constructed consisting of a glass tube approximately three feet in length and two inches in diameter. At the beginning of each test all water was evacuated from the tube. At a given signal, a pump was started causing water to rise in the tube. For the first ten inches, the rate of rise in the water column was observed by the *S*, and thereafter the tube was masked by a heavy tape. *S* was required to indicate when he believed the water column had reached a designated mark. Six trials were undertaken by each *S*, three at each of two fixed marks located at 20 and 28 inches respectively. *S* was not informed of the accuracy of his judgment. The score on this test was the average judgment in inches for the six trials. A low score indicated that the individual had underestimated the time required for the column to reach the mark.

2. *The Events Test*. This was the second test involving judgments of time. In this instance, six events of recent history were presented with the request that *S* indicate the approximate month and year on which each had occurred. These events were:

Supreme Court ruling on segregation.
Stalin's death.
The Kinsey Report.
The outbreak of the Korean War.
The McCarthy vs. Army hearings.
Eisenhower's operation.

This test was scored by first computing for each *S* the number of months from the present to the estimated date of each event. For every event the estimates of all *S*s were reduced to normalized *T* scores and then averaged for each individual. This procedure provided a single score for each *S* such that all event estimates were equally weighted regardless of the mean and *SD* of the estimates in terms of actual months. A low score indicated that the *S* tended to recall the events as nearer the present, and a high score that he estimated them as more remote.

TABLE 1
CORRELATIONS BETWEEN TUBE TEST, EVENTS TEST,
AND TARTAN TEST SCORED FOR *n* ACHIEVEMENT

	Events	Tartan (for <i>n</i> Achievement)
Tube Events	+.35	+.44 +.42

Note.— $p = .01$ at .46.
 $p = .05$ at .36.

3. *The Tartan Test*. This was the third test administered and consisted of 30 lithographic reproductions of Scottish tartans approximately 3 x 5 inches mounted on cards 8 x 11 inches in size. These reproductions were obtained from Robert Bain's *The Clans and Tartans of Scotland* (1) and varied widely with respect to predominant color, fineness of texture, complexity of design, degree of contrast, etc. The patterns were presented, mounted on a wall, to each *S*, with the request that he select the five most aesthetically pleasing and the five least aesthetically pleasing.

On the basis of a previous study (2), each tartan had been assigned a rank order for its correlation with achievement motivation as measured by McClelland (3). In the present study, each *S* was given a score consisting of the sum of the ranks of his five most preferred tartans minus the sum of the ranks of his five least preferred tartans. Thus a low or negative score indicated that *S* preferred tartans positively correlated with achievement motivation. The earlier study had shown that of the tartans occupying the first ten ranks, only the Ogilvie contained any significant amount of red. The rest, namely, the Campbell of Breadalbane, Elliot, Anderson, MacDonnel of Glengarry, MacPherson Hunting, Cameron of Erracht, Clergy, Sutherland Ancient, and Oliphant are uniformly somber and most show predominant blue and green color. On the other hand, the ten tartans yielding the largest negative correlations are strikingly different. Seven of the ten embody vivid red, namely, the Drummond, Hay, Sinclair, Brodie, Stewart, Ramsay, and Stewart of Appin. An eighth, the Barclay, contains vivid yellow, while of the remaining two, only one, the Cummings, has substantial blue or green and might be characterized as somber. In short,

this ranking of tartans progresses roughly from somber, or "ascetic," tartans to vivid "sensual" designs.

The correlations between the three primary measures of this study, the Tube Test, the Events Test, and the Tartan Test scored for *n* Achievement are given in Table 1. It will be seen that two of these correlations, namely those between the Events Test and the Tartan Test (scored for *n* Achievement) and between the Tube Test and the Tartan Test (scored for *n* Achievement) stand at the quite secure significance level of approximately .02. The third correlation is almost at the .05 level. Thus, there appears to be a fairly reliable pattern relating the tendency to recall past events as near the present, the tendency to anticipate future events before they occur, and a preference for that type of aesthetic asceticism known to correlate with high achievement motivation.

DISCUSSION

In an earlier study devoted to the analysis of Tartan preferences in relation to *n* Achievement, it was proposed that one of the prime qualities of the individual high in need Achievement is the desire to manipulate his environment and to view himself as the active agent. This consideration might lead us to conclude that preference for somber, passive

tartans merely represents a projected wish that his environment be *manipulanda*, not a *manipulator*. In this present study we wish to extend this line of thought in connection with attitudes toward time. We have shown here that persons who anticipate future conditions before they arrive also tend to recall past events as more recent than they really were. Such persons, it appears, wish to create in their psychological present a sort of "event density." We speculate that the future is already upon them while the past has not yet slipped away, and the universe that confronts them is therefore teeming with opportunities for manipulation and achievement. If this interpretation is correct, we may tentatively propose a dynamic triad, relating "harmonious" time attitudes, achievement motivation, and asceticism of aesthetic taste, which has historically found its manifestation in Puritanism.

REFERENCES

1. BAIN, ROBERT. *The clans and tartans of Scotland*. London: Collins, 1954.
2. KNAPP, R. H., *n* Achievement and aesthetic preference. In J. W. Atkinson (Ed.), *The assessment of human motives*. Princeton, N. J.: Van Nostrand, 1957.
3. MCCLELLAND, D. C., ATKINSON, J. W., CLARK, L. A., & LOWELL, E. L., *The achievement motive*. New York: Appleton-Century-Crofts, 1953.

Received October 11, 1957.

PATTERNS OF AUTHORITY AND AFFECTION IN TWO GENERATIONS¹

WANDA C. BRONSON, EDITH S. KATTEN, AND NORMAN LIVSON

Institute of Human Development,² University of California, Berkeley

ALTHOUGH there has been considerable speculation—and even consensus—concerning the differential roles of mothers and fathers in a given generation, little supporting empirical evidence has been presented. There has been perhaps even more speculation and less evidence regarding the specific nature of intergeneration changes in these roles. This report presents evidence concerning relationships among certain aspects of parental roles within the family. The general conceptualization that underlies the questions to be considered may be described briefly. In any given generation, parental roles are determined by a number of interacting factors. Among these are: a direct carry-over of learned roles from the preceding generation; a more complex and idiosyncratic intrapersonal mediation of early familial patterns; and contemporary social pressures and crises that force modification of certain existing parental roles. Insofar as a parent's definition of his role is affected by his childhood family pattern, this transmission is mediated both by his childhood and by his adult perceptions of his parents' behaviors and attitudes. The accuracy of his perceptions may vary widely. To take one example, he may as a child autistically attribute to his parents desired qualities which, in fact, are absent but, when grown and himself a parent, he may remember—and even exaggerate—these once-denied parental deficiencies. He then attempts, successfully or not, to "make up" to his child what his own childhood was lacking. Whatever the sources of his misperceptions, they will contribute to the changes and consistencies between his own and his parents' roles in the family.

Specifically, we consider two fundamental

¹ The authors are pleased to acknowledge the contributions to this investigation made by the staff of the Guidance Study. Our debt to all of those who have participated, over a span of 25 years, in this longitudinal study of child development and to its director, Jean W. Macfarlane, will be apparent throughout the present report.

² Formerly the Institute of Child Welfare.

aspects of the parent-child relationship: the giving of affection and the exercise of authority. These parameters, however abstract and lacking in specificity, appear nevertheless to be sufficient to outline the family pattern. Further, the intensity of the family interaction is evaluated by means of a measure of the degree of involvement that the parent feels both toward his own parents and toward his child. The data from the longitudinal investigation of a representative sample of American families, in which parents provided retrospective descriptions of their own childhood (in the early 1900's) and were observed in their behavior toward their own children (born in the late 1920's), permit the investigation of various aspects of the perception and transmission of parental attitudes in two generations. Four separate questions are considered: (a) how mothers and fathers of the earlier generation are remembered by their now-adult sons and daughters—the second generation parents; (b) in what ways and to what extent these retrospective assessments relate to the roles and behaviors adopted by these parents; (c) how these parents actually do behave to their children; and (d) how their children perceive these roles and behaviors, the accuracy of their perceptions, the nature and extent of their distortions. Lastly, by way of tying together these segmented analyses, our concluding section describes and illustrates the more characteristic patterns found in the second-generation families, drawing upon our direct and long-term observations of these parents and their children during the full span of development of the child into adulthood.

METHOD

Sample and Social Setting

Our data are drawn from the more than 200 families who participated in the Guidance Study of the Institute of Child Welfare. The purpose of this longitudinal study was the comprehensive investigation of physical, mental, and personality development in the children of the study families from birth through the 18th year.

The sample was closely representative of the population of children born in 1928-1929 in Berkeley, California. A more detailed description of this original sample and of the methods employed in collecting the data is presented elsewhere (Macfarlane, 1938).

The present study is based on 100 of these study families selected from the total group on the sole criterion that they offered over the longest time the most complete records with respect to parent-child relationships and the psychological development of the child. This subsample differs from the original group in exhibiting a somewhat higher socioeconomic level. Since the 100 study children are equally divided as to sex, we have subsamples of 50 boys and 50 girls for our investigation of the second-generation family. However, by making the reasonable assumption that the sex of the study child is unrelated to the parents' retrospective descriptions of their own parents (the grandparents of the study children), we may employ data from the full 100 families to specify fathers' and mothers' perceptions of their own parents.

To set the subcultural context in which these children were raised, it may be useful to cite certain gross parameters that describe the families included in the present study *at the time of the birth of the study child*. In 76% of the 100 families, both parents were native-born white; for 8%, neither parent was born in this country. In 66% of the families, marriages were between Protestants, in 20% between Catholics; the remainder, either mixed or of another denomination. The percentages of Protestant and native-born whites exceeded the then current norms for American urban communities. There were four Negro families in this group. The mean number of years of education of the parents was 11.9 for fathers and 11.5 for mothers (higher than the national average). Their mean ages at the time of the birth of the study child were, respectively, 32.7 and 29.0 years. Our sample of children contained 44 first-borns, although only 15 remained "only" children through the completion of the study period at age 18. Of the 100 families, 61% fell into the middle class and 37% into the lower class; in this respect, as well as in the occupational status of the father (both measured by Warner, Meeker, & Eells [1949] indices), the socioeconomic status of this sample was above average, although per capita income fell below the national mean. Although these families showed considerable improvement in their socioeconomic level through the study period (which included the depression and United States preparation for and participation in World War II), the impact of the depression on the family atmosphere through the early formative years of the child was often considerable.

Sources of Data

The full case records on which the ratings employed by this study are based include yearly interviews with the mother (half-yearly until age 8), occasional intensive interviews with the father, schoolteacher's reports, sociometric appraisals and, wherever possible, interviews with relatives. From age 6 on, half-yearly interviews were conducted with the child, and several projective techniques were administered on several

occasions. Since at least four interviewers contributed to the information available for each family, individual biases and blind spots should have had minimal effect on the consensus description. Although not all study families participated for the full 18 years of the study, all families treated in this report were available at least until the study child was 13 years old.

Rating Procedure

The basic data for this study consist of dichotomized ratings (High or Low) made by a single rater, highly familiar with the case records of all families, who abstracted and reviewed all material pertinent to each rating. Only when there was little or no information pertinent to a given rating was it omitted. When sufficient data were available, however equivocal their interpretation, a rating of High or Low was made: the ratings reflect differences in the level of intensity of a variable and not in the amount of information upon which they are based.

The study focuses upon the parents of the study families. Three broad areas are investigated: (a) the parents' *retrospective descriptions* of their relationship with their own parents; (b) the parents' *actual* behavior toward their children, and (c) the parents' behavior as *perceived* by their own children. In parents' descriptions of the grandparents and the children's perceptions of the parents, the ratings represent inferences concerning the actual beliefs held by the respondents which, of course, do not necessarily correspond to their stated attitudes. The ratings of parental behavior evaluate the actual treatment accorded by the parents to their child. Although the child's own reports of parental behavior would have been potentially of great value in assigning these ratings, they were excluded from consideration in order to prevent contamination of this set of ratings by those which evaluate the child's perceptions. All three of these broad areas were rated for the two main variables of the study—authority and affection. The third variable, involvement, was rated only for the parents' descriptions of the grandparents and for their actual behavior toward the child.

Description of Variables

The variables used in this study have very strong connotative loadings so that it becomes particularly important to avoid attributing to them—and thus to our results—any surplus meanings that exceed the following descriptions.

Authority. In the parents' descriptions of the grandparents, as well as in the children's reports, authority was rated as the degree to which the parental figure was *perceived* by the respondent as attempting to control his behavior. In rating the parents' *actual* behavior, their intent to discipline the child, and their implementation of this intent by methods deemed to be recognizable and effective for the average child were used as the criteria for judgment. The underlying motivation and emotional quality of the control and its effectiveness for the particular child were irrelevant to the rating. For example, the mere presence of highly socialized behavior in the child was not taken as evidence, per se, for a

High rating of authority in either parent. By way of underlining the qualitative differences subsumed by the same rating, we may note that the feeling tone accompanying the exercise of authority ranged in the sample from harshness to reasonableness and consistency.

Affection. In the parents' descriptions of the grandparents and in the children's reports, the respondent's perception of a quality of warmth, acceptance, and love was the basis for the rating. In the judgment of the actual degree of affection shown by each parent, the rater sought behavioral manifestations over time, however subtle, which would have been sufficiently overt to be recognizable by a child. Thus, for example, if the child is denied any *visible* sign of parental love, whether through personal inhibition or subcultural proscription, affection is rated as Low. Again, as in the case of authority, the underlying motivation for affectionate behavior did not affect the rating, and many qualitative differences were subsumed under the same ratings: Highs on affection, for example, range from "smothering" to genuine affection and support.

Involvement. The intensity of the emotional relationship served as the basis for the rating of involvement in both the parents' descriptions of the grandparents and in their actual behavior toward the children. In the former, the rating focused on the parent's memory of his own childhood relationships with his parents; it largely reflects the degree to which he feels that his parents affected his past and present attitudes and behavior. In the latter, the rating is of the emotional investment of the parent in his child—the extent to which he regards the child as an extension of himself and attempts to "live through" him. As in the case of authority and affection, these ratings are concerned merely with the level of involvement, not with the "goodness" of the relationship.

Reliability

Two measures of reliability were obtained using two different randomly drawn subsamples of 25 families. To test the self-consistency of the original rater, one of the subsamples was re-rated by her after an interval of about three years. The original order in which the families had been rated was reversed, thus preventing an inflation of reliability due to practice effects. To assess the reproducibility of the ratings by a rater far less familiar with the total case records, an interrater reliability measure was obtained on the second 25-family subsample. Agreement indices were based only on instances in which the data had been judged on both of the two rating occasions to permit a rating. The over-all ratable percentage, thus defined, for the intrarater subsample was 81%; the over-all agreement index on these ratable instances, 87%. In the interrater subsample, the over-all ratable percentage was 66%; the percentage agreement, 82%. In Table 1 we report the ratable percentage and the percentage agreement on both reliability checks for each category and for each variable. Although, as would be expected, the ratable percentages were lowest for ratings concerning grandparents (for whom the data were least complete), the agreement values are uniformly and adequately high for all categories and for the three

TABLE 1
SUMMARY OF INTRATER AND INTERRATER
AGREEMENT ON RATINGS BY VARIABLE AND BY
CATEGORY^a

	Percentage Ratable ^b		Percentage Agreement	
	Intra	Inter	Intra	Inter
By Category				
Mother re MGM	84	51	97	84
Mother re MGF	69	43	96	88
Father re PGM	61	45	89	82
Father re PGF	53	43	82	84
Mother re Child	100	95	84	84
Father re Child	92	88	83	80
Child re Mother	100	96	80	81
Child re Father	96	88	83	77
By Variable				
Authority	81	65	87	83
Affection	83	70	86	83
Involvement	77	64	89	78

^a Two different subsamples of 25 families were used for each of the two separate evaluations of rating reliability. Abbreviations used are: MGM (Maternal Grandmother), MGF (Maternal Grandfather), PGM (Paternal Grandmother), and PGF (Paternal Grandfather).

^b Refers to the percentage of the maximum possible rating instances where a rating of "High" or "Low" was available for both rating occasions; the agreement indices are based on these instances only.

RESULTS AND DISCUSSION

Since all our ratings are dichotomous, associations among sets of ratings were set up as 2×2 contingency tables, the significance of which was estimated by the chi-square test for correlated groups (always with a correction for continuity) except in those instances when an exact (Fisher) computation of probability was permitted by published tables or required by a too small minimum theoretical expected frequency (<5). The N for each comparison is the number of cases that proved ratable (by the original rater) for *both* of the sets of ratings being compared. Contingency coefficients (C) were computed to evaluate the extent of the significant or near-significant relationships. Since all these coefficients are based on 2×2 tables, the reported values are comparable; the maximum possible value is always .71.

The reported percentages of High ratings for each variable are based on the maximum data available, i.e., all the ratable instances for the given set of ratings. However, the tests of significance of the percentage differences between any two ratings, which were made by McNemar's test of difference of correlated proportions (using the continuity correction),

were based, necessarily, on the common N for the particular comparison. Whenever available tables allowed, the exact probability of the observed difference was computed from the cumulative binomial expansion. Implicit in our use of the varying common N for a given association or difference comparison is the unavoidable assumption that, in each instance, these subsamples are representative of the full sample of 100 families. No systematic source of bias is apparent to us in this procedure.

The presentation of the findings is organized in terms of the questions posed at the outset. For a study so frankly exploratory, we feel it advisable to report even slight trends that do not meet customary levels of statistical significance. Furthermore, since we are dealing with closely interrelated aspects of family life, several trends—though each insignificant in itself—may, considered together, point to a meaningful generalization. For these reasons, all results permitting the rejection of the null hypothesis at the .15 level are reported.

Fathers' and Mothers' Perceptions of the Grandparents

The retrospective descriptions of first generation family patterns are summarized in Fig. 1. It is immediately apparent that grandfathers and grandmothers are perceived very

differently by fathers and mothers in our sample. Significantly more grandmothers than grandfathers are remembered by fathers as highly affectionate, and there is a tendency for more of them to be reported as having been a strong source of authority. Mothers do not recall the grandparents' behavior as having been different; further, they differ from fathers in more often remembering grandfathers as highly affectionate. Regarding the parents' emotional involvement, more fathers than mothers show High Involvement with the grandmother; also, fathers are involved significantly more with the grandmother than with the grandfather. Whether these differences in fathers' and mothers' reports reflect actual sex differential treatment by the grandparents or represent distortions originating in either childhood or in adulthood cannot be ascertained from such retrospective data. Be this as it may, we find that the grandmother is significantly more figural than the grandfather in the fathers' view of their childhoods—a generalization not applicable to the mothers' descriptions.

Further delineation of first generation familial patterns is made possible by considering the interrelationships among Authority, Affection, and Involvement within the reports of fathers and mothers. (A total of 15 comparisons for each parent is generated by this analysis.) The summary in Table 2, which presents all obtained significant and near-significant relationships indicates that both parents tend to remember affection as a general quality of the home; it was rare for the quality to be High in one grandparent and not in the other. Authority, on the other hand, is characteristically found High in only one of the grandparents. Furthermore, when it is the grandmother who is a strong authority she is less often regarded by both mothers and fathers as highly affectionate. This mutual exclusiveness of authority and affection also occurs in the description of the grandfather—strongly so for mothers ($C = .35$) and only tentatively for fathers ($C = .19$). Apparently, whether in retrospect or in fact, the strong exercise of authority by either parent tended to preclude the presence of a High level of affection in that parent.

The correlates of involvement also differ for fathers and mothers: Fathers report High

Differences in Parents' Descriptions of Grandparents

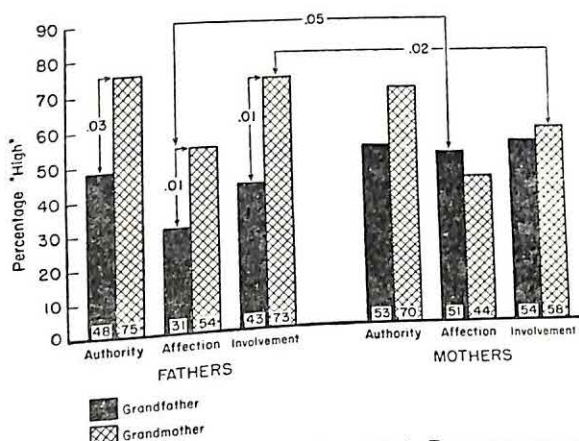


FIG. 1. DIFFERENCES IN PARENTS' RETROSPECTIVE DESCRIPTIONS OF GRANDPARENTS

(All possible percentage differences between fathers and mothers and between grandfathers and grandmothers on each of the variables, considered separately, were tested for significance. No comparisons between the three variables were made. Only probabilities of .15 or less are reported.)

TABLE 2

SUMMARY OF SIGNIFICANT RELATIONSHIPS BETWEEN AUTHORITY, AFFECTION, AND INVOLVEMENT IN GRANDFATHERS (GF) AND GRANDMOTHERS (GM) AS PERCEIVED BY FATHERS AND MOTHERS

Perceived by	Comparison	Direction	<i>p</i>	<i>C</i> ^a	<i>N</i>
Both	GF Authority × GM Authority	Negative	.002	.38	63 (Fathers)
			.02	.31	53 (Mothers)
	GF Affection × GM Affection	Positive	.07	.22	70 (Fathers)
			.06	.23	60 (Mothers)
	GF Authority × GF Affection	Negative	.14	.19	60 (Fathers)
			.01	.35	54 (Mothers)
Fathers	GM Authority × GM Affection	Negative	.05	.28	65 (Fathers)
			.06	.24	61 (Mothers)
	GF Involvement × GM Involvement	Negative	.11	.20	61
		Positive	.03	.28	56
	GF Authority × GF Involvement	Negative	.09	.23	54
		Positive	.02	.30	59
Mothers	GM Affection × GM Involvement	Positive	.02	.29	69

^a The maximum possible value for these contingency coefficients is .71.

involvement with grandfather when he is remembered as High in authority; for mothers there is a positive relation between involvement and remembered affection in relation to both grandmother and grandfather. For fathers, strong involvement with one parent tends to exclude strong involvement with the other; this does not hold for mothers. Maternal authority seems to have special meaning for mothers: the exercise of a high level of authority by the grandmother tends to preclude a report of High involvement with the grandfather. This finding suggests that the abdication—in fact or fantasy—by the male of his traditional power role resulted in his daughter's alienation from him.

Interrelationships Between Parents' Perceptions and their Behavior Toward the Child

There is very little support for our expectation that the actual child-rearing practices of the parents would be systematically related to their descriptions of their own parents. Out of 72 possible comparisons (mothers' and fathers' characterizations of their own parents on each of the three variables compared with the three aspects of their actual behavior towards their own sons and daughters) only two trends emerge: (a) Mothers more often exercise strong authority in the home when they remember their own mother as having done so. Although this relationship has low significance for both sons ($p = .12$) and daughters ($p = .10$) considered separately, an over-

all evaluation for boys and girls taken together yields a significant relationship ($p < .02$, $C = .31$). (b) Fathers show a similar tendency to emulate the same-sex parent but, for them, it is in the area of affection that the relationship prevails. Again, the relationships singly are of low significance (sons, $p = .12$; daughters, $p = .07$), but are strong for the combined groups ($p < .01$, $C = .29$).

The dearth of significant relationships found between the parents' reports of their parents and their behavior toward their own children seems to imply little direct transmission of roles from generation to generation. We should remember, however, that the parents' reports are retrospective—over a considerable time span—and are therefore highly susceptible to distortion. When we further consider that evaluations of parental figures are very likely emotionally "loaded," their veridicality becomes even more suspect, particularly so when we consider that the respondents have themselves assumed the roles that they are describing. Clearly, the situation may invite distortions in the service of resolving persisting childhood conflicts or perhaps providing self-justification for current inadequacies, real or felt.

The Parents' Actual Practices

A summary description of the ratings of parents' actual relationships toward their children is presented in Fig. 2. Except for authority, the roles of fathers and mothers are

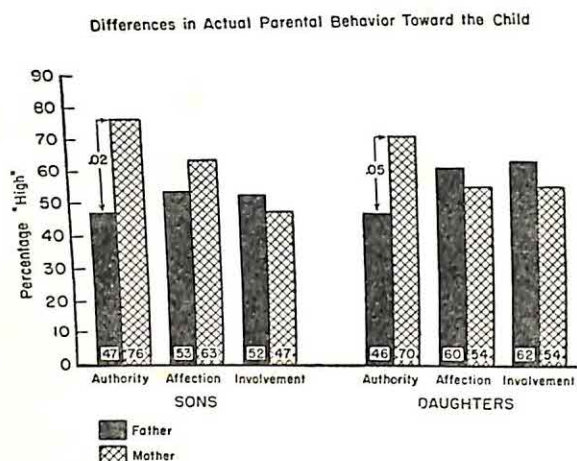


FIG. 2. DIFFERENCES IN ACTUAL PARENTAL BEHAVIOR TOWARD THE CHILD

(All possible percentage differences between sons and daughters and between fathers and mothers on each of the variables, considered separately, were tested for significance. No comparisons between the three variables were made. Only probabilities of .15 or less are reported.)

similar: for boys and girls alike, both parents are equally likely to offer High affection or to become strongly involved. The finding that significantly more mothers than fathers are rated as High in authority both towards their sons and their daughters conforms to the frequently offered generalization concerning the primary role of the mother in the disciplining of the child. However, contrary to our expectation, neither mothers nor fathers were seen as giving differential treatment in any area as a function of the sex of the child.

To what extent are these three facets of parental behavior interrelated? Table 3 (in which the total possible number of comparisons is 15 each for boys and girls) shows that the only relationship reported in common for both sons and daughters lies in the behavior of the mothers: a mother highly affectionate to her children will also be highly involved with them. For sons only, such positive relationship between affection and involvement obtains in fathers' behavior. These three very strong relationships lead to the not very surprising inference that a parent's involvement with his child is expressed, at least in part, in affectionate behavior towards him. By contrast, then, the absence of any such relationship in fathers' behavior toward their daughters may be taken to suggest that in many instances an

additional factor—perhaps some prohibition against overt affection when the father feels strong ties towards his daughter—renders the two less compatible.

The relationships unique to daughters all point to a certain mutual exclusiveness in the behavior of the two parents. When one parent is either strongly authoritative or very involved the other is not. Further, in two instances, involvement from one parent shows a tendency to relate negatively to some characteristic of the other parent: these relationships occur between mothers' involvement and fathers' authority and between fathers' involvement and mothers' affection. This trend does not appear for sons; in fact, the only relationship unique to them is a positive one which indicates that they tend to receive the same degree of affection from both parents.

The Child's Perception of Parental Behavior

What are the child's perceptions of the parental practices? Figure 3 portrays the children's reports in the areas of authority and affection. On the whole, these descriptions coincide with the interviewers' ratings of actual parental behavior. Both sons and daughters perceive mothers, more often than fathers, as exercising strong authority in the home, and they see no difference in the frequency with which fathers and mothers are highly affectionate. However, although we have noted above that sons and daughters do not receive differential treatment from either mother or father, there is a slight tendency for more sons than daughters to perceive their fathers in a strong authority role. This implication that sons seek to augment the fathers' authority is further supported by the finding that, when the sons' perceptions disagree with the interviewers' ratings, they err in the direction of perceiving a High level of authority when a Low level was adjudged as actually present ($p = .06$). Since daughters show an opposite tendency (not in itself significant), there is a significant sex difference in the direction of distortion of the fathers' authority ($p = .05$). There is no evidence for systematic distortions in the perceptions of mothers' authority nor in the perception of affection from either parent.

This general absence of distortion, systematic or not, is reflected in the over-all accuracy

TABLE 3

SUMMARY OF SIGNIFICANT RELATIONSHIPS BETWEEN AUTHORITY, AFFECTION, AND INVOLVEMENT IN THE ACTUAL PRACTICES OF FATHERS (F) AND MOTHERS (M) TOWARDS SONS AND DAUGHTERS

Behavior toward	Comparison	Direction	<i>p</i>	<i>C</i> ^a	<i>N</i>
Both	M Affection × M Involvement	Positive	.001	.46	49 (Sons)
			.001	.43	50 (Daughters)
Sons	F Affection × M Affection	Positive	.08	.25	46
	F Affection × F Involvement	Positive	.001	.47	47
Daughters	F Authority × M Authority	Negative	.05	.27	48
	F Authority × M Involvement	Negative	.09	.24	48
	F Involvement × M Affection	Negative	.13	.22	47
	F Involvement × M Involvement	Negative	.04	.30	47

^a The maximum possible value for these contingency coefficients is .71.

of the children's view of their parents' behavior toward them. The relationship between the interviewers' ratings of authority and affection of fathers and the children's perceptions of these characteristics is in all instances significant beyond the .001 level (the contingency coefficients range from .44 to .61). For mothers, agreement on authority also reaches the .001 level for both sexes (boys, $C = .60$; girls, $C = .45$) but for affection the relationships are somewhat weaker (boys, $p < .01$, $C = .39$; girls, $p = .04$, $C = .29$). Expressed in terms of the percentage of agreement between the interviewer and the child, the fathers' authority level is correctly perceived by 89% of the sons and 85% of the daughters; while on affection both sexes are correct in 80% of the cases. Mothers' authority tends to be perceived more correctly by sons (94%) than daughters (82%): this difference approaches significance ($p = .08$). There is no sex difference in the accuracy of perceiving mothers' affection; the values are 74% for boys and 67% for girls. Both sexes are more accurate regarding mother's authority than her affection (sons, $p = .04$; daughters, $p = .12$). None of the children show any difference in the accuracy with which they perceive the mother and father.

Our evidence does not support an assumption that perceptiveness in a child is a general quality. We find that accuracy in the perception of authority is unrelated to an accurate perception of affection and, with but a single exception, no relationship obtains between the accuracy with which the father and the mother are perceived. The single instance is that sons who correctly evaluate the level of affection in one parent tend to do so for the other ($p = .07$, $C = .26$). Implicit in our treatment of the

discrepancies between the child's perception and the rating of the actual behavior of the parent as "distortions" and "misperceptions" is the assumption that such differences derive primarily from the child's attempt to perceive the world more in accordance with his needs. We recognize that both the unreliability of the ratings and our failure to take into account qualitative differences in the parents' expression of authority and affection that could modify their perception by the child no doubt contribute to these discrepancies and therefore raise doubts concerning their treatment as indices of accuracy. However, the generally very high level of correspondence between the two sets of ratings argues that such extrinsic factors have minimal weight. Furthermore, the very fact that systematic differences do occur—both in direction of distortion and in level of accuracy—seems to support the usefulness of regarding discrepancies as distortions.

Proceeding to the interrelationships between parental authority and affection as perceived by the children we find that of the six comparisons possible for boys and for girls, three reach significance. None of these coincide with relationships found in the ratings of actual parental behavior (see Table 3). One of the relationships is common to both sons and daughters: children apparently cannot regard their mothers as simultaneously High in authority and affection. Mothers' affection and authority are negatively related, at the .01 level, for sons ($C = .31$) and daughters ($C = .36$). (This same negative relationship was found in both parents' retrospective descriptions of grandmothers.) Since this relationship does not appear in the actual behavior of mothers and since, as we have already shown, affection is more often misperceived than is authority by

Differences in Childrens' Perceptions of Parental Roles

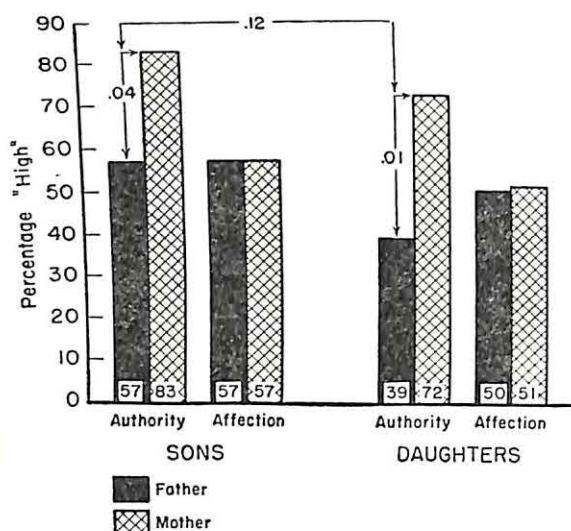


FIG. 3. DIFFERENCES IN CHILDREN'S PERCEPTIONS OF PARENTAL ROLES

(All possible percentage differences between sons and daughters and between fathers and mothers on each of the variables, considered separately, were tested for significance. No comparisons between the two variables were made. Only probabilities of .15 or less are reported.)

both boys and girls, we may suggest that it derives from a tendency of children to refuse to credit affection to mothers who assume strong authority roles. To test this hypothesis, the 22 children who correctly perceived the level of their mothers' authority but who were inaccurate in their judgment of affection were examined to determine the direction of their error. For 16 children the distortion involved lowering the level of affection—a systematic misperception significant at the .05 level. Since we found no such trend in the total group, this result lends support to the hypothesis.

Two other relationships, perceived only by sons, are too low to merit extensive analysis. There is a slight tendency for sons to deny that both parents have equal authority: perceived levels of authority for mothers and fathers are negatively related ($p = .12$, $C = .23$). Also, sons tend to see a positive relationship between fathers' affection and mothers' authority ($p = .11$, $C = .24$).

A final area of analysis is the comparison of the children's and their parents' perceptions of the authority and affection of parental figures. (The data underlying these comparisons are

drawn from Figures 1 and 3.) The confounding of two sources of variation in these data, however, prevent their unequivocal interpretation; mothers' and fathers' descriptions are retrospective evaluations, made in adulthood, of their childhood some 25 years ago. Hence, any difference between these perceptions and those of the children may reflect actual cultural change in parental behavior as well as the effect of distortions related to the passage of time. Only two differences emerge. Comparing mothers' and daughters' perceptions we find that paternal authority is more frequently remembered as High by mothers than it is so seen by daughters ($p = .02$). The father-son comparisons show only that sons, more often than fathers, report a highly affectionate male parent ($p = .04$).

Family Patterns

Thus far our treatment of these data has been essentially a univariate one. In this concluding section we consider the patterning of authority, affection, and involvement exhibited in the actual practices of our study families. The examination of the relative frequencies of these patterns for mothers and fathers—while of course not independent of the previously reported parental differences—does serve to illustrate more clearly the interactions among aspects of parental behavior. (In this analysis we deal only with the 93 of the 100 families for whom complete data were available.) From Table 4 we see that 50% of mothers and 42% of fathers fall into two of the eight possible configurations. This striking

TABLE 4
FREQUENCY OF PATTERNS OF AUTHORITY, AFFECTION, AND INVOLVEMENT^a

Mother				Father			
Au- thority	Affec- tion	In- volve- ment	Frequency	Au- thority	Affec- tion	In- volve- ment	Frequency
High	Low	Low	24	Low	High	High	21
High	High	High	23	High	High	High	18
Low	High	High	15	Low	Low	Low	14
High	High	Low	14	High	Low	Low	13
High	Low	High	8	Low	High	Low	9
Low	Low	Low	7	High	Low	High	8
Low	High	Low	2	Low	Low	High	6
Low	Low	High	0	High	High	Low	4

^a Based on the ratings of actual behavior in the 93 families for which complete data were available on mothers' and fathers' behavior toward the child.

homogeneity can be interpreted as demonstrating a considerable uniformity in cultural standards for parental behavior. One of the two modal patterns (High, High, High) is common to mothers and fathers: the role of a highly authoritative and affectionate parent who has a strong emotional investment in the child is apparently equally accessible to both sexes. The other most frequent pattern for mothers is of a woman who, while clearly assuming responsibility for disciplining the child, gives him little warmth or involvement (High, Low, Low). An exact inversion of this pattern characterizes the fathers' other modal role: a man who is very warm and involved but unwilling—or unable—to exercise strong parental authority (Low, High, High).

What are the characteristic-combined mother-father patterns within a single family? Our data suggest that family patterning is nonrandom: an over-all comparison between actual frequencies of family patterns with the expected values generated by the hypothesis of random pairing of the single-parent patterns yields a chi-square value of 67 ($p = .02$). This finding may imply the existence of some interaction relating to authority, affection, and involvement, either in the selection of partners in marriage or in the assignment of roles within the family after marriage. Of the two most frequent family patterns, one originates from the four most expected combinations, i.e., from one of the two modal patterns of each parent. Six of our families consist of a High, Low, Low mother and a Low, High, High father. This particular pattern calls to mind the "typical" American family in which mothers "wear the pants," focused upon by commentators on our culture and frequently exploited in our entertainment media. The other modal pattern which also occurs in six families—Low, High, High for both mother and father—is not drawn from among these four. It seems curious that we find such a relatively high frequency of a family constellation in which neither parent assumes clear responsibility for the socialization of the child. We may speculate whether these families are representatives of the theme of "permissiveness" and "self-regulation" in child-rearing that appeared to reach its peak during the period in which these families raised their children (Vincent, 1951; Wolfenstein, 1951).

Our integration, thus far, has dealt only with interactions among the variables as defined and rated in this study. As a re-examination of our definitions will indicate, these variables are dimensions abstracted from the full complexity of family relationships which, when rated dichotomously and summarized over time, could hardly reproduce the subtleties and comprehensiveness of the original observations. The relationships that we report may be attenuated—to an unknown extent—by our failure (unavoidable in a normative study) to use more of the potential information. In order to portray our data more closely to their original form, we present a case description of a somewhat extreme representative of one of our two most frequent family patterns (mother: High, Low, Low; father: Low, High, High).

The mother of this boy was an overpowering woman with extraordinary energy and confidence. Her average weight over the years was 200 pounds. She said of herself: "I guess mother is the iron hand in this family." Interviewers described her as aggressive, volatile, and completely unperceptive although cheerful and well-organized. She used rigid discipline with the total self-confidence that whatever she did was best for everybody, that there are certain standards to conform to, with no room for discussion. Her stated belief was that the boy would never amount to anything if he were not pushed and corrected; she was frank in wanting him tied to her. She over-protected him from the outside world but deflated and hemmed him in at home, almost swallowed up his life. Her enveloping domination was, in her own mind, the expression of her complete devotion to the boy, but, as she showed neither tenderness nor demonstrativeness, by no one else could it be interpreted as affection. She was incapable of seeing the effect of her drives on her introverted child, and never faced the existence of his serious personality handicaps. She treated him as she had treated his older sisters, as objects to be bettered by yielding to her will, not as individuals to be understood and identified with.

The father, a slight, gentle man, was described by interviewers as a friendly person, mellow and quiet, with a great deal of underlying strength and stamina, who attained a high level of competence in his profession. He hated unpleasantness, avoided disciplining, but defended the boy when he could. In the early years the father took over the "mother" role of bathing the boy, was the one who kissed him, and was both easy and companionable with him. The father and son, the "henpecked" ones, had much in common and expressed mutual admiration in their quiet, reserved ways. The father tried to make time to talk with his son, to be a source of support for him, and what pleased the father most was that the boy looked like him, acted like him, and was following in his footsteps.

The son, a frail boy, passively submitted to the rigid

domination of his mother, stating, "Everybody at my house has to mind my mother," and saying that he would not dare to talk back to her, "even my papa doesn't." Her overpowering personality, in conjunction with the loved father's passive role with her, resulted, for this boy, in his becoming somber, shy, and inhibited. He was described as serious, worrisome, and very conscientious. He looked and acted defeated, was tense and pathetic. At 8 years, the boy said he would like to be 18 "so I won't have to mind my mother" but, reflecting, added, "No, 20 is safer, then I could go away and be my own boss." But at 17, he still had to struggle to keep back the tears when he told of the mother's domineering behavior and said that he had no relationship with her except one of irritation.

The father-son relationship had a very different quality. At 7 years, the boy said, "I kiss my father because I like him the best," and expressed the sort of feeling that a younger boy more usually has for his mother. At 9 years, the boy was unhappy that the father no longer had time to give him his bath. Over the years he spoke with intense admiration of his father who, to him, was "the greatest living person." From 14 years on, the boy said he would be afraid to get married, "It might turn out like my father's," and his hostility to the mother seemed more focused and articulate. He patterned his behavior after the father's and was glad he was a boy so he could be a chemist like the father who was "an inspiration to him." In his last interview, at 17 years, he said, "I'm closest to my father; I think he knows how I feel mostly."

This case is illustrative, not typical; for each of the remaining five families sharing this pattern there is much that is unique both in the expression of the pattern and its effect on the child. We find among them a boy who is described as "happy, cheerful and unaggressive," and another characterized as "the sort who would be happiest in the Hitler Youth." Despite this diversity, a given family pattern may be regarded as setting the common situation with which the child must cope while the qualitative differences in the expression of the pattern, as well as the resources that the child possesses, combine to determine the nature of his solution to the situation.

SUMMARY

Cumulative case history material on 100 families, drawn from an 18-year intensive investigation of physical, mental, and personality development, was rated on the three variables of Authority, Affection, and Involvement. These ratings were made with respect to: (a) each parent's retrospective descriptions of his own parents; (b) the consensus of objective descriptions of the parents' behavior toward their child; and (c) the children's perceptions of each of their parents' behavior.

The reliability of these dichotomous ratings (High or Low) was evaluated in two separate subsamples of 25 families each: over-all intrarater agreement was 87%; interrater agreement was 82%.

The data were specifically examined with reference to the following four questions: (a) How are the roles of fathers and mothers in the earlier generation remembered by their now-grown sons and daughters? (b) What is the nature of the transmission of parental roles from one generation to the next? (c) What are the actual behaviors and attitudes of the parents of the study families toward their children? And, (d) how do these children perceive their parents' behavior? Among the main findings relating to each of these questions are the following: (a) In the parents' retrospective descriptions, the maternal figure is described by both mothers and fathers as having been significantly more often a strong authority figure than was the father and, when in this role, to have been relatively unaffected towards her children. (b) With respect to the transmission of parental roles, there is little *direct* carry-over from parents' descriptions of their parents to their behavior toward their own children. (c) In actual behavior, the mother is significantly more often a strong source of authority than is the father; there are no mother-father differences on affection or involvement. We find no differences on any of the three variables between the behavior of either parent toward his son or daughter. (d) On the whole, children are very accurate in their perceptions of parental behavior although this accuracy does not appear to be a generalized ability—perceptual accuracy in the several areas is not interrelated. The patterns of the parents' behavior on the three variables are presented and the modal family patterns are discussed.

REFERENCES

- MACFARLANE, JEAN W. Studies in child guidance: I. Methodology of data collection and organization. *Monogr. Soc. Res. Child Develpm.*, 1938, **3**, 6, 1-254.
- VINCENT, C. E. Trends in infant care ideas. *Child Develpm.*, 1951, **22**, 199-209.
- WARNER, E. L., MEEKER, MARCHIA, & EELLS, K. *Social class in America*. Chicago: Sci. Res. Associates, 1949.
- WOLFENSTEIN, MARTHA. The emergence of fun morality. *J. soc. Issues*, 1951, **7**, 15-25.

Received November 27, 1957.

FACTORS IN THE MOTOR BEHAVIOR OF FUNCTIONAL PSYCHOTICS¹

P. H. VENABLES

Medical Research Council Social Psychiatry Research Unit, Institute of Psychiatry, London, England

IN AN earlier study (5), data derived from the performance of functional psychotics on a five-choice repetitive task were found to be explicable in terms of a tentative theory that hypothesized an exaggerated tendency for schizophrenics in relation to normals to develop reactive inhibition. In the present study, the aim was to determine whether schizophrenics would continue to show this hypothesized greater decrement in response rate in a task where higher response rates were possible. A further aim was to determine whether reactive inhibition is the central factor responsible for the decrement in performance and subsequent reminiscence during rest periods.

Previously, the task involved five stimulus lights paired with five response keys. When the subject (*S*) responded to a light by pressing the key paired with it, he extinguished the light, and the next light was presented only after *S* had pressed a central sixth key. In the present study, a modification made higher response rates possible. At the same time he extinguished the stimulus light by pressing the response key paired with it, *S* was instrumental in setting up the next light without the necessity of using an intermediary sixth key.

In the earlier study, where response rates ranged from 30–50 responses per minute for schizophrenic *Ss*, the pattern of their performance was an increase in speed for an initial 10-minute period, followed after a rest by a decrease in the second 10 minutes.

METHOD

Subjects. The *Ss* were 50 male chronic schizophrenics having a mean age of 38.9, *SD* 8.7, years. They had been hospitalized for a mean of 7.1, *SD* 5.7, years, with a minimum of 2 years. For comparison, a group of 14 male patients suffering from endogenous depression was tested; they had a mean age of 55.1, *SD* 6.7, years

and had been hospitalized for a mean stay of 5.4, *SD* 7.0, years.

Eighteen members of the artisan and nursing staff of the same hospital acted as normal controls; their mean age was 37.2, *SD* 14.0, years. The 50 schizophrenic *Ss* were selected from a larger population of *Ss* having the requisite characteristics of age, length of stay, and diagnosis, according to their positions on a scale of activity-withdrawal (6) to form equal sized active and withdrawn subgroups.

Apparatus. A display panel contained five lights, set two inches apart in a horizontal row on a plane 15 degrees from the vertical and perpendicular to the line of sight of the average *S*. Immediately in front of each light and in the horizontal plane were five keys, 5 in. long and 1½ in. wide, operating microswitches. The whole of the apparatus except for the keys, which were unpainted aluminum, was painted mat black. The lights had intensities of 60 ft. candles, and the background illumination had an intensity of 1.7 ft. candles. In operation, *S* pressed the key corresponding to the light which was lit, thus extinguishing it and replacing it by another light, which was then responded to, and so on. The sequence of lights was random over a series of 50, with the restriction that no light ever appeared twice in succession. The sequence was permuted at the end of each 10-minute run. Responses were recorded on an event recorder.

Procedure. *Ss* were told that the task was to see how quick they were and that any questions they asked would be answered at the end of the experiment.

Each *S* worked for 10 minutes, followed by a rest of 5 minutes, after which he worked for a further 10 minutes. Ten days later, he worked for a further 10 minutes, had 5 minutes' rest, and then worked for 3 minutes.

RESULTS

Each *S's* responses in each work period were summarized by linear regression slope constants. The constant representing slope was designated by "a" and that representing initial level by "b." The other main variable considered in the analyses was the amount of reminiscence, "R," calculated by subtracting the number of responses made on the last minute of a work period from the number made on the first minute of the succeeding period.

As seen in Table 1 and Fig. 1, there were differences between the levels at which each group performed at each stage of the experiment. *F* ratios and *p* values in the last two columns of Table 1 indicate that for each 10-

¹ This work was carried out in Springfield Hospital, London, by kind permission of the Medical Superintendent, H. C. Beccle. The author wishes to acknowledge the help and advice given by his colleague, J. Tizard.

TABLE 1
MEAN INITIAL RESPONSE SPEEDS FOR EACH GROUP
FOR EACH TEN-MINUTE WORK PERIOD

	Schizophrenics		Depressives	Normals	<i>F</i>	<i>P</i>
	Active	Withdrawn				
1b	81.08	70.78	78.00	94.15	8.25	<.001
2b	91.82	80.72	87.80	105.10	12.04	<.001
3b	93.70	80.17	92.30	106.73	6.81	<.001

minute period these differences are highly significant.

The first part of Table 2 gives values for the mean regression slopes for each group in each of the three 10-minute work periods. Analyses of variance show no evidence of significant differences between groups.

Examination of Fig. 1 gives the impression that the degree of slope exhibited by each group may be a function of the speed at which the group works. Correlation coefficients were therefore calculated between initial speed and slope over all groups for the three 10-minute work periods. They were $-.450$, $-.233$, and $-.256$, respectively. Because these correlations are all significant, it was decided to carry out analyses of covariance to determine significance levels between groups and to provide adjusted mean slope constants for each group in each period with the variance attributable to initial speed held constant.

The second part of Table 2 shows these results. While none of the differences between groups was significant, the analyses of covariance indicated mean slope values more in line with expectations. The schizophrenics showed a greater decrement than normals and depressives in all periods, in contrast to the results with unadjusted data, where, for instance, withdrawn schizophrenics showed significantly less diminution of response speed than normals in the first five minutes of work ($t = 2.30$, $p > .05$).

Further examination of Fig. 1 suggests that in the first and third work periods the active schizophrenics and normal controls decreased in rate in the early part of the period, showing a later increase. Examination of the raw data for these groups showed that in the first work period there was adequate evidence for a fall and a subsequent rise in speed with a point of inflexion for the majority

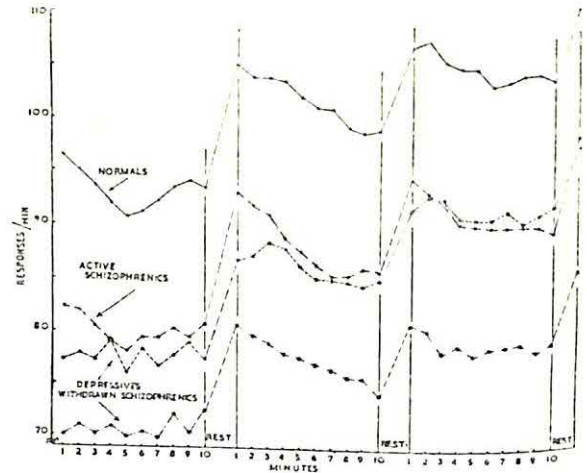


FIG. 1. MEAN RESPONSE SPEED PER MINUTE FOR ACTIVE AND WITHDRAWN SCHIZOPHRENIC, DEPRESSIVE, AND NORMAL CONTROL GROUPS

of Ss between the fifth and sixth minute. Separate linear regression slopes were therefore calculated for the first and second five minutes' performance separately. The data for the third 10-minute performance indicate that the mean rise in response speed later in the period was not shown in the majority of Ss and that a single point of inflexion could not be found for those Ss showing the late rise in rate. Separate regression lines were therefore not calculated.

Results of analyses of variance and covariance for the data from the first and second five-minute performance are shown in the lower part of Table 2. As with the analysis of the whole of the first 10-minute period, none of the results were significant although covariance adjustment tended to produce values more in line with the expectation of greatest decrement in schizophrenics.

Table 3 shows the mean values of reminiscence for each group at the end of each 10-minute work period. In no case was there any evidence of significant differences.

Several associations between the variables of slope and reminiscence are predictable on the basis of the suggestion by Venables and Tizard (5), following the Mowrer-Miller hypothesis, that decrease in rate of response is due to the accumulation of reactive inhibition and that reminiscence is due to spontaneous dissipation of this inhibition.

If the same factor of reactive inhibition is responsible for the decrease in response rate during a work period and for the amount of

TABLE 2
MEAN AND ADJUSTED MEAN REGRESSION
SLOPES FOR ALL PERIODS

	Schizophrenics		Depres- sives	Nor- mals	F	P
	Active	With- drawn				
1a	-0.40	+0.22	-0.04	-0.37	0.409	NS
2a	-1.68	-1.30	-0.69	-1.35	2.12	NS
3a	-1.13	-0.44	-0.23	-0.69	0.937	NS
1a adj.	-0.31	-0.40	-0.21	+0.52	0.815	NS
2a adj.	-1.59	-1.53	-0.81	-1.11	0.670	NS
3a adj.	-1.05	-0.95	-0.17	-0.25	0.994	NS
1a 1st half	-1.33	-0.06	-0.24	-1.46	4.28	<.01
1a 2nd half	+0.16	+0.53	+0.09	+0.58	0.61	NS
1a 1st half adj.	-1.25	-0.80	-0.44	-0.51	0.998	NS
1a 2nd half adj.	+0.17	+0.14	+0.05	+1.00	2.50	ap- prox. .1

TABLE 3
MEAN REMINISCENCE VALUE FOR EACH GROUP AT
THE END OF EACH TEN-MINUTE WORK PERIOD

	Schizophrenics		Depres- sives	Normals	F	P
	Active	With- drawn				
1R	12.20	8.32	8.43	11.16	1.02	NS
2R	8.68	6.72	6.35	7.94	0.22	NS
3R	8.84	6.80	6.93	7.05	0.42	NS

reminiscence shown during a subsequent rest period, there should be a negative correlation between slope and reminiscence scores. Similarly, slopes in different work periods and reminiscence scores in different work periods should correlate positively. In view of the lack of significant difference between groups on these variables, the data for all groups were combined and intercorrelations calculated. All intercorrelations save that between 1a and 2a, significant at the .01 level, were significant at the .001 level, thus confirming the predictions.

The general rise in the size of correlations between variables later in the task confirms the work of Reynolds (4). In order to demonstrate the existence of a common factor responsible for these interrelations, a centroid factor analysis was carried out with two iterations. Three factors were extracted, although Factors II and III were of borderline significance. Factor I, which accounts for 54% of the total variance, suggests a common determinant as the basis of the reminiscence and slope measurements. Factors II and III do not appear susceptible to interpretation.

TABLE 4
INTERCORRELATIONS BETWEEN REMINISCENCE AND
SLOPE MEASURES FOR THE TOTAL GROUP AND
UNROTATED LOADINGS FROM FACTOR ANALYSIS

	1a	2a	3a	1R	2R	3R	I°	II°	III°
1a		.300	.454	-.430	-.380	-.454	-.554	.249	-.215
2a			.681	-.664	-.641	-.558	-.819	-.203	.277
3a				-.422	-.629	-.691	-.821	-.143	-.204
1R					.363	.523	.704	-.308	-.372
2R						.477	.714	.336	.078
3R							.769	.124	.159

Correlations were carried out between reminiscence and slope measures and age and length of stay in hospital for each group of Ss; in no case was a significant correlation found. Lack of correlation between slope of performance and length of stay does not run counter to the difference in amount of slope shown by short- and long-stay groups in the previous study (5) because all Ss used here fell in the long-stay (>2 years) category.

DISCUSSION

Two main trends emerge from the results. First, about half the variance of the data concerned with decremental slope and reminiscence in performance seems attributable on the basis of factor analysis to a single factor. Second, there is no evidence of any difference between the amounts of decrement in performance or subsequent reminiscence shown by schizophrenic, depressive, or normal groups throughout the experiment; when differences in initial speed of performance are not allowed to influence the results, however, there is a consistent tendency in all periods for schizophrenics to show a greater decrement in response rate than normals.

If the factor responsible for decrement in performance and subsequent reminiscence is reactive inhibition, it is necessary to examine how far the hypothesis that schizophrenics are characterized by an exaggerated tendency to build up reactive inhibition is weakened by the lack of significant differences in slopes and reminiscences between the schizophrenic and the normal or depressive groups.

While it cannot be proved that Factor I, resulting from the intercorrelations between slope and reminiscence measures, is a factor of reactive inhibition, features of the performance in this experiment show considerable similarity to those in experiments on inverted alphabet printing by Kimble (2) and on the

pursuit rotor by Ammons (1). In these cases, reactive inhibition I_R or its equivalent of "temporary work decrement" (D_{wt}) were fruitfully hypothesized as appropriate intervening variables. The consistencies among these studies give reactive inhibition a marked degree of support.

The main alternative hypothesis to account for decrement in schizophrenic, if not normal, performance is that of decline in motivation. If this were the factor responsible, it seems hardly likely that the change in level of motivation in the first five-minute rest period would produce a greater reminiscence in all groups than the second rest period of 10 days. The finding of a smaller amount of reminiscence in the second and third rest periods is, however, in line with explanations in terms of reactive inhibition. Kimble and Shatel (3), for example, showed that reactive inhibition diminishes in later trials. Similarly, the negative and subsequently positive slopes of performance in the first and second five minutes generally replicates the work of Ammons (1), who showed that D_{wt} (his equivalent of I_R) reaches a maximum at from three to eight minutes in work on the pursuit rotor and thereafter remains stable or declines. If, then, in the present experiment, I_R is at its maximum in the fifth minute of the first work period, it is to be expected that subsequent performance will be influenced by other factors, most likely those concerned with learning.

It is suggested that the nonsignificant differences in development of I_R in this experiment, compared with the significant difference found earlier, may be explained by an hypothesis, capable of subsequent testing, concerned with the speed at which I_R is dissipated in various classes of Ss. If the task is sufficiently easy for responses to be made rapidly, then the responses may be so close together that the I_R generated by a single response is not dissipated before the next response is made; therefore, I_R accumulates. At high speeds, even normal Ss will be unable to dissipate I_R , and the difference in amounts accumulated by different groups of Ss is a function of differential amounts of I_R generated by each response. At lower speeds, as in the previous experiment and a pilot experiment on normal Ss which preceded it, normal and depressive Ss are able to dissipate what-

ever amount of I_R is generated by one response before the next occurs. However, schizophrenic Ss, in whom it is hypothesized that I_R dissipates more slowly, are unable to dissipate wholly the I_R which they generate before the next response occurs, so I_R accumulates.

While some of the features concerning changes in performance thus seem to be capable of hypothetical explanation, the reason for the large significant differences in actual rate of performance shown by the different groups remains unexplained. No hint is given by the data in terms of, for instance, diminution of a general level of motivation due to hospitalization. Length of stay correlates with initial speed level only to the extent of $-.22$ ($p > .05$) for the schizophrenic groups and $-.19$ ($p > .05$) for the depressive group. Correlations between age and initial speed are also insignificant, $-.01$ and $-.06$, respectively.

SUMMARY

In a repetitive task, no significant differences were found between regression slopes showing decrement of response speed or reminiscence values in the performance of schizophrenic, depressive, and normal Ss. Factor analysis and subsequent discussion suggested that the factor responsible for response decrement and reminiscence might be considered as reactive inhibition. An hypothesis is proposed to explain the differences found between the results for the present and a previous similar study.

REFERENCES

1. AMMONS, R. B. Acquisition of motor skill: II. Rotary pursuit performance with continuous practice before and after a single rest. *J. exp. Psychol.*, 1947, **37**, 393-411.
2. KIMBLE, G. A. An experimental test of a two-factor theory of inhibition. *J. exp. Psychol.*, 1949, **37**, 15-23.
3. KIMBLE, G. A., & SHATEL, R. B. The relationship between two kinds of inhibition and the amount of practice. *J. exp. Psychol.*, 1952, **44**, 355-359.
4. REYNOLDS, B. The effects of learning on the predictability of psychomotor performance. *J. exp. Psychol.*, 1952, **44**, 189-198.
5. VENABLES, P. H., & TIZARD, J. Performance of functional psychotics on a repetitive task. *J. abnorm. soc. Psychol.*, 1956, **52**, 23-26.
6. VENABLES, P. H. A short scale for rating "Activity-Withdrawal" in schizophrenics. *J. Ment. Sci.*, 1957, **103**, 197-199.

Received October 10, 1957.

TWO TYPES OF CONDITIONING IN PSYCHOTICS AND NORMALS¹

N. O'CONNOR AND K. RAWNSLEY

Medical Research Council Social Psychiatry Research Unit, Institute of Psychiatry, London, England

THE acquisition of conditioned responses in man has recently been studied extensively in its connection with anxiety (15, 16). However, no definitive study of the effect of psychosis on the acquisition of CRs has yet been done. Among those reports available, that of Peters and Murphree (11) is representative of the commonly held view that psychotics condition less readily than normals. In this respect, these authors consider their results a confirmation of Gantt's (4) view that conditionability is an indicator of clinical condition and of Pavlov's view of schizophrenia as "pathological inertia."

Other research has produced conflicting findings. Spence and Taylor (14) could not demonstrate that psychotics condition more readily than neurotics or normals. Welch and Kubis (18) put schizophrenics at the extreme inhibitory end of their scale of conditionability and inhibition. Similarly, Howe (5), using a PGR technique in normals, anxiety neurotics, and schizophrenics, predicted that schizophrenics have a low drive level with respect to anxiety in anticipation of electric shock and hence condition less readily than normals. By the criterion of the size of response, however, the results were not statistically significant. Franks (2), working with Ss who were in the main other than chronic long-stay cases, found no particular tendency for psychotics to condition more or less readily than normals.

Because few systematic studies of the conditionability of schizophrenics as such have been attempted, this experiment was designed to investigate the performance of two subgroups, paranoid and nonparanoid chronic schizophrenics. Previous studies of the learning of chronic schizophrenics (9, 10) suggested that differences might be expected between these two groups.

Although many different diagnostic subgroups must have been included under the heading of "psychosis" in the studies cited,

¹ This work was carried out in Banstead Hospital, Surrey, by kind permission of the Physician Superintendent, E. P. H. Charlton. The authors are indebted to their colleagues and especially N. C. Surya for many valuable comments and suggestions.

most used one or other of the two usual forms of conditioning, the conditioned eye-blink or the conditioned GSR. Various other procedures have been developed in other connections, based in many cases on Skinner's (13) work with rats. Verplanck's (16) studies are perhaps the best known of these, although Lindsay and Skinner (7) have applied their own techniques most consistently to the problem of the responses of chronic psychotics. Interest here was in the comparison of a commonly used conditioning technique, the eye-blink, and a verbal conditioning technique having something in common with Verplanck's but being much simplified for use with psychotic patients. This is a technique of verbal conditioning used in Russian studies by Luria (8) and reported in detail by Povorinski (12).

METHOD

Subjects

Three samples were compared: 20 normal males aged 28 to 59 years, 20 nonparanoid chronic schizophrenic males aged 33 to 55 years, and 20 chronic paranoid schizophrenics aged 35 to 55 years. The respective mean ages of the groups were 39.40, *SD* 12.51; 41.45, *SD* 5.84; and 47.20, *SD* 4.94. Each of the patient groups was selected according to Bleuler's diagnostic criteria as judged from clinical record and interview, and an additional criterion of more than two years' hospital residence was also fulfilled in each case. The normal group was composed of male nurses randomly selected from a group of approximately the same age range as the patient groups.

Apparatus

The apparatus and criteria used in the eyelid conditioning procedure were similar to those used by Franks (3). It differed in one way: a modified light source² was attached to S's head. Head movements thus did not result in variations of illumination on the photoelectric cell. This considerably reduced the possibility of random head movements being recorded as eye-blinks. The nature of the apparatus also allowed the photoelectric cell to be directed at the center of the eye of each S, irrespective of the shape of the head.

The apparatus used in the verbal conditioning procedure consisted of a circuit in which a secret switch operated a light signal and a recording clock. A hand pressure on another switch closed the circuit, extinguished the light, and stopped the clock. The latency

² This apparatus was designed and built by Eric Sayer.

of response could thus be measured. The circuit and the technique were evolved by Luria (8) and others and described in Povorinski's (12) handbook.

Procedure

Eyelid conditioning. The Ss in the eyelid conditioning situation were seated in a darkened quiet room in a comfortable chair, and a light beam and photoelectric cell were adjusted to point at the center of the right eye while the eye was gazing at a fixation light placed on a level with it. The patient was then told to look at the fixation point of light whenever this light came on and to remain looking at it until it had gone off again. Otherwise, head and arm movements were not restricted. The fixation light was switched on 3 to 7 seconds before the stimulus was given and switched off about 5 seconds afterwards.

The auditory threshold was then determined at a tone of 1100 cps, and the intensity of the CS was adjusted to a level of 65 db above the threshold. Following Franks' procedure, three conditioned stimuli (tones), three unconditioned stimuli (air puffs), and a further three CS were applied. An eye-blink response to two of the latter three CS was taken to indicate the presence of pseudo-conditioning, and the S was rejected from the experiment in this case. The US was an air puff at a pressure equivalent to 65 mm. of mercury and lasting 500 msec. The CS was a pure tone delivered to both ears at the frequency and intensity mentioned above and for a duration of 800 msec. The CS and the US overlapped for 450 msec.

Forty-eight conditioning trials followed the initial trials. Thirty were combined air puffs and tones. Eighteen were tone alone (test trials). The sequence of reinforced and unreinforced stimuli was randomized, and the interval between stimuli was also randomized within limits of 20 sec. and 40 sec. The sequence was the same for each S.

An eye-blink was scored whenever the pen of the recording ammeter recorded a response of 150% of the maximum amplitude recorded during the alerting period between the switching on of the fixation light and the sounding of the CS. In this way the possibility of mistaking random eye movements for CRs was minimized. The response was recorded as a CR when it occurred within 1.25 sec. of the sounding of the CS. No minimum time criterion was laid down. No record was accepted if six of the first ten reinforced signals did not result in an eye-blink. The retest reliability of eye-blink scores reported by Franks (2) was 0.52. Interrater consistency when judging records, however, can be even less satisfactory, and great care needs to be taken to define criteria as precisely as possible with this measure.

Verbal conditioning. The verbal conditioning technique followed immediately. S was told nothing more but seated in a chair in front of a desk on which there was a small light (12) and a hand-press lever on which he was told to rest his preferred hand. He was then told that he would soon see what was happening and that there was no need for further explanation. A few seconds later the light was switched on. Two seconds were allowed to elapse, and then E said, "Press." When S had pressed on his lever and the light had gone off, E said "Good." Times of presentation of stimuli

were randomized between 40 and 70 sec. in the first conditioning series. Trials were continued until S pressed as soon as the light appeared and without instruction continued to respond in this way for 10 consecutive trials.

At this stage, an interrupting task consisting of counting aloud backwards from 100 for a minute was introduced. Following this, the latency of response to the signal was noted, and the procedure continued until it had returned to its previous norm. A new signal was then introduced, consisting of the word "light" in the place of the light itself. Comparative latencies to the two signals could thus be compared.

An Introversion Questionnaire (1) was given to all Ss.

RESULTS

The results can be considered in two parts, eyelid acquisition scores and verbal conditioning scores. Table 1 gives the means and standard deviations by groups for both scores.

The eyelid acquisition score is the number of CRs in the 18 test trials. The verbal conditioning score is the number of trials required for the first stable acquisition of a CR, i.e., one followed by 10 CRs without the UCS. When an analysis of variance was carried out, F ratios calculated for the above results were found to be 7.63 in the case of verbal conditioning ($p = \text{near } .001$) and 1.78 in the case of eyelid conditioning (not significant). Analysis of covariance was then calculated on the verbal conditioning scores to eliminate the effects of age, which differed somewhat from group to group. This was done only with verbal scores because the eyelid scores did not differ significantly. This recalculation reduced F to 6.53 ($p < .01$). The mean acquisition score was not materially altered, remaining at 15.19 for the mean number of trials necessary in order to condition the paranoid Ss.

Mean extinction scores were available in

TABLE 1
CONDITIONING SCORES BY GROUPS

	Eyelid Acquisition Score		Verbal Conditioning Score	
	Mean	SD	Mean	SD
Normal ($N = 20$)	4.80	4.27	6.35	5.98
Chronic paranoid schizophrenic ($N = 20$)	4.45	3.04	15.20	10.71
Chronic nonparanoid schizophrenic ($N = 20$)	6.55	3.62	6.75	6.01

TABLE 2
MEAN LATENCIES OF RESPONSE IN SECONDS

Group	Before Interrupting Task ^a	After Interrupting Task	Significance Level of Increase	Mean of last 3 Responses to Light	Mean of First 3 Responses to Word "Light"	Significance Level of Change
Normals	0.59	0.97	.02	0.69	0.49	.01
Nonparanoid schizophrenics	1.50	2.07	.02	1.85	1.25	.01
Paranoid schizophrenics	1.67	1.90	NS	1.57	1.46	NS

^a That is, counting backwards for one minute.

the case of the eyelid conditioning. These were 2.05, *SD* 2.82, *CRs* in the series for the normals, 1.75, *SD* 1.99, for the paranoid schizophrenics, and 3.15, *SD* 2.43, for the nonparanoid schizophrenics ($F = 3.55$; $p = .05$). The t (2.6) between paranoid and normal *Ss* was significant ($p = .02$). Other differences were not significant.

The result of interrupting a series of responses by a mental task and of replacing a light signal by such a verbal signal as the word "light" can be seen from Table 2.

When the interrupting task was introduced, latencies of *CRs* increased, although the general trend prior to the interruption was towards a decrease in latency with repetition. The increase was greatest in the normal control and nonparanoid groups, where it reached the .02 level of significance, and least in the paranoid group, where it was not statistically significant.

The change to a verbal signal disrupted the response of both the psychotic groups and to a lesser extent the normal group. Eleven of the 20 paranoids and 12 of the 20 nonparanoids needed reconditioning as compared with 5 of the 20 normals. The average number of trials needed was 1.3 and 0.6 for the paranoids and nonparanoids and 0.25 for the controls.

A comparison of the mean of the last three direct signals with the first three verbal signals shows (Table 2) that the introduction of a verbal signal resulted in significantly decreasing latency of response for normal and nonparanoid groups but not for the paranoids. There was no significant difference between the last response time in response to the light

signal and the first response time to the word "light" for any group.

The correlations between introversion scores and speed of conditioning were not statistically significant when calculated for each subgroup for both eyelid conditioning and verbal conditioning. When those in the paranoid schizophrenic group who conditioned slowly with the verbal procedure, i.e., who took more than 10 trials, were compared with those who took fewer trials to condition, Fisher's Exact Probability Test showed no correlation between the two measures.

DISCUSSION

Previous investigators, except for Franks (2), claimed a difference in conditionability due to clinical condition. This difference has not been borne out here for the eyelid conditioning procedure. In this respect, Franks' results are confirmed. With verbal conditioning, however, clear differences between the groups were apparent in the mean scores, even taking account of age differences. The slow conditioning of the paranoid group may point to a delay due to the use of words. On the other hand, it may mean a delay in the acquisition of a conditioned motor response. In this case, it would be similar to the finding of relatively slow learning in the case of paranoid schizophrenics reported elsewhere (9, 10). If a delay due to the use of words was involved, the results may resemble those of Vinogradov and Reiser (17). Referring to the phenomenon of "elective irradiation" in the secondary signalling system, a phenomenon similar to that described by Krasnagorski (6), Vinogradov and Reiser note that disturbances of this type of irradiation may characterize paranoid schizophrenia.

The tendency to poor verbal conditioning was characteristic of about half the paranoid schizophrenics. Thus, if a cutoff point is fixed at 12 responses, about one *SD* above the mean for the nonparanoid schizophrenics and the normals, then only one member of these two groups exceeds this score, whereas nine of the paranoids do so.

The verbal conditioning procedure involves motor activity which may presumably be delayed or suppressed entirely by the *S*. In this respect, the readiness with which conditioning occurs may depend in part on such

factors as attention, concentration and, more remotely, on the extent to which *S* cooperates in the experimental situation. From the psychiatric standpoint, the reaction to a face-to-face situation and communication through words in the verbal conditioning experiment may conceivably account for the relatively delayed conditioning of the paranoid group. It is suggested that paranoid attitudes engender negativism and resistance to demands made by others—in this case the investigator—and that these factors impose a delay in the establishment of the conditioned response.

The results of an interrupting task showed a difference between paranoids and the other two groups. Similarly, the change to a verbal stimulus reduced the latency of the normals and nonparanoids but not that of the paranoids. In each of these two instances, as in the first verbal conditioning trials, the paranoid group behaved differently from the other two. It remains to decide whether the difference is a difference of motor response, a difference of cooperation, or a difference connected with responses to verbal material. In the opinion of the authors, the explanation is likely to be in terms of a disturbance in the relationship of speech and action. It can be argued, for example, that because the paranoid group shows no significant increase following an interrupting task, there is no evidence of a lack of cooperativeness. Also, there is no evidence from the latency times that the paranoids are significantly different from the nonparanoids in speed of motor response.

For these reasons, although the mode of operation of any verbal-motor peculiarity in chronic paranoid schizophrenics remains to be determined, the material presented favors the suggested explanation. Alternative explanations in terms of differences in cooperativeness are made more unlikely by the behavior of the 40 chronic psychotic patients who, with only one or two exceptions, sat alone in darkness for 30 minutes, wearing an uncomfortable apparatus and receiving during this time irregular blasts of air in the center of the right eye.

SUMMARY

A normal group ($N = 20$), a chronic paranoid schizophrenic group ($N = 20$), and a

chronic nonparanoid schizophrenic group ($N = 20$), all males, were compared for conditionability on eyelid conditioned response and a verbal conditioning technique. No statistically significant differences were observed between the groups for eyelid conditioning, but the paranoid group took significantly longer to condition on the verbal procedure. The results are different from some previous findings with eyelid conditioning. These and other reported findings with the verbal techniques may indicate a verbal process involvement in the paranoid group which is not typical of chronic nonparanoid schizophrenics, although other explanations in terms of cooperativeness are feasible.

REFERENCES

1. EYSENCK, H. J. A short questionnaire for the measurement of two dimensions of personality. *J. appl. Psychol.*, 1958, **42**, 14-17.
2. FRANKS, C. N. An experimental study of conditioning as related to mental abnormality. Unpublished doctoral thesis, Univer. of London, 1954.
3. FRANKS, C. N. Conditioning and personality: A study of normal and neurotic subjects. *J. abnorm. soc. Psychol.*, 1956, **52**, 143-150.
4. GANTT, W. H. Measures of susceptibility. *Amer. J. Psychiat.*, 1943, **99**, 839-849.
5. HOWE, E. S. A theoretical and experimental investigation of the conditioned galvanic skin response in normals, anxiety states, and schizophrenics. Unpublished doctoral thesis, Univer. of London, 1952.
6. KRASNAGORSKI, N. K. Physiology of cerebral activity in children as a new subject of pediatric investigation. *Amer. J. Dis. Child.*, 1933, **46**, 473-494.
7. LINDSLEY, O. R., & SKINNER, B. F. New techniques of analysis of psychotic behavior. *O.N.R. Ann. Tech. Rep.* 3, 1955-56, Proj. Nos. N5-ori-07662 & Nonr-1866(18).
8. LURIA, A. R. Some problems of the higher nervous activity of normal and abnormal children. In A. R. Luria (Ed.), *Problems of the higher nervous activity of normal and abnormal children*. Moscow: Acad. of Pedagog. Sciences of the R.S.F.S.R., 1956.
9. O'CONNOR, N. Reminiscence and work decrement in catatonic and paranoid schizophrenics. *Brit. J. med. Psychol.*, 1957, **30**, 188-193.
10. O'CONNOR, N., HERON, A., & CARSTAIRS, G. M. Work performance of chronic schizophrenics. *Occup. Psychol.*, 1956, **30**, 153-164.
11. PETERS, H. N., & MURPHREE, O. D. The conditioned reflex in the chronic schizophrenic. *J. clin. Psychol.*, 1954, **10**, 126-130.
12. POVORINSKI, U. A. *Metodika issledovaniya dvigatel-*

- nik ooslovnik refleksov na rechevom podkreplenii.* Leningrad: Biblioteka Prakticheskovo Vrach, Medgiz. Leningradskoye Otdeleniye, 1954.
13. SKINNER, B. F. *The behavior of organisms.* New York: Appleton-Century, 1938.
 14. SPENCE, K. W., & TAYLOR, JANET A. Anxiety and strength of the UCS as determinants of the amount of eyelid conditioning. *J. exp. Psychol.*, 1951, **42**, 183-188.
 15. SPENCE, K. W., & TAYLOR, JANET A. The relation of conditioned response strength to anxiety in normal, neurotic, and psychotic subjects. *J. exp. Psychol.*, 1953, **45**, 265-272.
 16. VERPLANCK, W. S. The operant conditioning of human behavior. *Psychol. Bull.*, 1956, **53**, 70-83.
 17. VINOGRADOV, N. V., & REISER, L. A. Experimental study of the interaction of the first and second signal systems in schizophrenics as related to the clinical course of the disease. *Zh. vyssh. nervn. Deiatel'*, 1953, **3**(1), 77-91. (*Psychol. Abst.*, 1954, **28**.)
 18. WELCH, L., & KUBIS, J. Conditioned psycho-galvanic response in states of pathological anxiety. *J. ment. nerv. Dis.*, 1947, **105**, 373-381.

Received October 17, 1957.

SOME CORRELATES OF BELIEFS IN THE MALEVOLENCE AND BENEVOLENCE OF SUPERNATURAL BEINGS: A CROSS-SOCIETAL STUDY¹

WILLIAM W. LAMBERT, LEIGH MINTURN TRIANDIS, AND MARGERY WOLF

Cornell University

A STUDY of some of the social psychological functions of "primitive" religious beliefs, this paper tests cross-societally some hypotheses about how general anticipations of pain develop in children, and the relation of these to aspects of the formal belief systems of a society. The major hypothesis was that beliefs in the *malevolence* of the supernatural world reflect punitive practices in infant and child rearing, while beliefs in the *benevolence* of the supernatural world reflect nurturant practices in infant and child training. The research program in which the present study is included investigated antecedents of aggression by means both of direct field study (8) (not reported here) and of ratings derived from ethnographic reports of 62 societies.

METHOD

The societies rated for the present paper include a wide range of geographic settings. Thirty-one of Murdock's² 61 world culture areas are represented by at least one society. Our sources were necessarily limited to those affording adequate descriptions of child training practices. Most of the ethnographic sources were selected from the bibliography of Heinicke and Whiting (5). The geographical distribution of our "sample" is shown in Table 1.³

¹ The present study was facilitated by grants from the Social Science Research Council and from the Ford Foundation (Cornell-Harvard-Yale Socialization Study) as well as by the facilities of the Center for Advanced Study in the Behavioral Sciences. The authors are indebted for data and suggestions to Irvin Child (and his collaborators), Elisabeth Lambert, Wallace E. Lambert, Charles Morris, Morris E. Opler, and John W. M. Whiting. We are also indebted to William and Corinne Nydegger, Florence Rosenberg, and to members of two Seminars at Cornell University who helped develop the scales and did many of the ratings.

² Murdock, G. P., *World Ethnographic Sample*. (Mimeographed paper)

³ It might be argued that only one culture from each of Murdock's culture areas should be represented in our "sample." We considered this, but the plan was dropped when we noted that in all areas there was at least one culture from each area which had a different

None of the tables of results includes all these cultures because in each case the available sources provided inadequate information for some of the ratings. Table 2 lists most of the societies studied. We attempted ratings that would reflect the *general* benevolence and aggressiveness of the supernatural belief system, on the assumption that there is some basic coherence in the "traits" underlying the various representations that the gods and spirits may take within the belief systems of a culture. The raters had therefore to consider the ethnographer's report of all situations in adult life in which the supernatural appears in any form; and specific beliefs in supernatural intervention in formal religion, ritual, witchcraft, and sorcery were thus all of some importance in making the judgments.

The ratings on socialization practices and on beliefs about gods were made independently, the former under the direction of Irvin Child at Yale University and the latter by members of a research seminar under the direction of Lambert and Triandis at Cornell. A recent analysis of the Yale data has been presented in two papers by Child et al. (3) and Barry et al. (1).

Socialization Scales⁴

Independent ratings of socialization measures were made by two judges on a 7-point scale, the sum of the two ratings being the score.

Infancy. Ratings were made on several variables covering the first year of a child's life and as long thereafter as the treatment of the infant remains approxi-

pattern of relationship on the deities scale and infant pain scale from the others, and in four of the nine areas there were no repetitions of patterns on the two variables.

⁴ In this study, only the infancy and childhood ratings related to boys are used. Ratings for girls were also available. Ratings for boys were chosen because another study (in preparation) on the role of women for these same cultures found that women very rarely develop high status in those power roles that may tend to determine some changes in culture. Their child training, where it differs, may have different functions in cultural integration. This problem has been set aside for later analysis.

TABLE 1
GEOGRAPHICAL DISTRIBUTION OF
SOCIETIES IN SAMPLE

Polynesia	Micronesia	Melanesia	Indonesia
6	3	7	7
Africa	Eurasia	North Amer.	South Amer.
11	5	17	6

TABLE 2
RELATION BETWEEN ABSENCE OF PAIN FROM
NURTURING AGENT IN INFANCY AND
PROPERTIES OF THE SUPERNATURALS

High Pain (Low absence of pain)	Low Pain (High absence of pain)
Supernaturals—mainly aggressive	
Alor	Andaman
Aymara	Bena
BaVenda	Lepcha
Chagga	Lesu
Chirricahua Apache	Manus
Dahomey	Wogeo
Kurtachi	Yagua
Kwakiutl	
Kwoma	
Maori	
Navaho	
Ojibwa	
Lovedo	
Siriono	
Tenetehara	
Tepoztlan	
Thonga	
Supernaturals—mainly benevolent	
Arapesh	Ashanti
Chamorro	Chenchu
Klamath	Cheyenne
Ontong-Java	Comanche
Ovimbundu	Fiji
Puka Puka	Hopi
Tallensi	Papago
	Samoa
	Teton
	Tikopia
	Winnegago
	Zuni

NOTE.— $p = .05$. 95% confidence limits for the relative frequency (29/14 = .68) = .51 — .81 (4, pp. 66-69)

mately constant. The nine scales used were: *protection from environmental discomforts, absence of pain inflicted by nurturant agent, over-all indulgence, diffusion of nurturance, display of affection, consistency of drive reduction, immediacy of drive reduction, degree of drive reduction, and constancy of presence of nurturant agent*. The scale on over-all indulgence covers all the data on which the other infancy scales were rated, except diffusion of nurturance, in addition to other general statements on the topic made by the ethnographer. In regard to the scale of environmental discomforts, raters considered the extent to which these were not experienced, were usually prevented, or were quickly eliminated. Pain inflicted by nurturant agent included such things as cold baths, depilation, and so on, as well as physical punishment. In considering diffusion of nurturance,

raters judged the degree to which nurturance is shared by others than the mother, who was thus used as reference point.

Childhood. Ratings were made for the period between infancy and puberty (roughly 5-12 years). Scales for six behavior areas were rated: *nurturance, responsibility, self-reliance, achievement, obedience, and general independence*. For each of the scales, ratings were made on the basis of "positive training" (both reward for presence and punishment for absence of the behavior), punishment for nonperformance only, punishment for performance, and frequency of performance. In scoring "positive training," raters considered frequency, degree, consistency, and immediacy of reward for performance and of punishment for nonperformance. Examples of rewards are adult approval, approval of contemporaries, status gain, basic drive reduction, material gain, and anxiety avoidance. In scoring punishment, raters considered severity and frequency. Examples of punishments are corporal punishment, disapproval, deprivation (of freedom, food, etc.), threats, and natural consequences (such as curtailed freedom, food, etc.).

Scales on Properties of the Deities⁵

A god or spirit was defined as any supernatural being who was capable in principle of responding to the actions of tribal members. The definition was designed to include diverse kinds of gods and spirits and to exclude such impersonal life forces as *mana*. In scoring aggressive behavior by the gods, judges were instructed to consider the frequency with which gods were considered responsible for such occurrences as famines, plagues, weather disturbances, personal mishaps, etc. Examples of benevolent behavior ascribed to the gods include protection from enemies, granting personal favors, curing, growing good crops, good hunting, etc. The bad things and good things that happen to people in a primitive society are much the same everywhere—sickness, death, love, birth, good hunting, or good crops cover most of the instances, and in terms of frequency the weight probably lies with the first four. Our procedure amounted to asking what proportion of these good or bad "things" are referred to the supernatural.

Two ratings were made of benevolence and aggressiveness with regard both to frequency of benevolent or aggressive action and to intensity of the modal action. The frequency scales were stated as the proportion of all the acts of the gods which were aggressive or benevolent. Any act could be considered as benevolent or aggressive, as both benevolent and aggressive, or as neither benevolent nor aggressive. The intensity measures were an over-all rating of the intensity of the gods' benevolent or aggressive actions. A culture was termed either "mainly aggressive" or "mainly benevolent" on

⁵ Roberts et al. provide a relationship that gives further meaning to our interpretation of this scale. They related this scale to the presence or absence of games of chance in a society and found that where the supernaturals were judged to be aggressive more than half the time the people did not play chance games, but that where the supernaturals were benevolent more than half the time games of chance were played. (Roberts, J., Bush, R., & Arth, M. J. *Mastery in games: A cross-cultural study*. [Mimeo.]

the basis of the arithmetic relation between the two frequency scales, with the intensity scales being used to determine ties. Seven-point scales were used. If the ratings obtained independently by two raters diverged by more than two points, the raters conferred on the evidence involved in their judgments, and where agreement was not reached the score was omitted. Ratings with a disagreement of two points or less were averaged. In 85% of all the cases, the independent original ratings differed from each other by two points or less.

We report the following socialization data, knowing that the various measures of both infancy and childhood treatment are statistically and sometimes definitionally related. Our argument outlines one way of explaining some of the common variance.

RESULTS

Infancy

A clear relationship between absence of pain from nurturing agent in infancy and properties of the supernaturals is shown in Table 2. This relationship is significant at the .05 level.

The relationship of other infancy variables to the aggressive or benevolent properties of the deities is displayed in Table 3. Although

TABLE 3
RELATIONSHIPS BETWEEN INFANCY TREATMENT
VARIABLES AND THE AGGRESSIVENESS AND
BENEVOLENCE OF THE SUPERNATURALS

Infancy Training Variable	Supernaturals Predominantly:		Significance Tests ^a
	Aggressive	Benevolent	
High protection from environmental discomforts	36% (8/22) ^b	61% (11/18)	NS $p = < .05$; c.l. for 29/43 = .51-.81
High Absence of Pain Inflicted by nurturant agent	29% (7/24)	63% (12/19)	
High over-all Indulgence	39% (11/28)	60% (12/20)	NS
High diffusion of Nurturance	44% (12/27)	68% (13/19)	NS
High Display of Affection	33% (8/24)	50% (10/20)	NS
High Consistency of Drive Reduction	39% (9/23)	53% (10/19)	NS
High Immediacy of Drive Reduction	35% (8/23)	47% (9/19)	NS
High Degree of Drive Reduction	37% (10/27)	44% (8/18)	NS
High constancy of Presence of Nurturing Agent	52% (14/27)	37% (7/19)	NS

^a We report significance level by χ^2 test, usually as .05 level with no further refinement. We also report 95% confidence limits for either the proportion of "successful" cases or of "unsuccessful" cases.

^b The figures in parentheses show the numbers on which the percentages are based. The numerator is the number of tribes showing the characteristic at the left out of the number of tribes classified as having either predominantly aggressive or benevolent gods (the denominator).

none of these additional relationships emerges as significant, all of them are similar in direction in the sense that the children in societies with predominantly aggressive gods and spirits are less cosseted. The only exception to this generalization is that the nurturant agent tends to be more constantly present in the societies with aggressive deities. Even when only those societies are considered where little pain is inflicted by the nurturant agent, there is still a tendency for the nurturant agent to be present less often in benevolent cultures. This result may perhaps mean that what is done to the child by caretakers is more closely related to the properties of the deities than is the mere fact of the presence of the caretaker.

If we combine the diffusion of nurturance ratings with the over-all indulgence ratings, the pattern of low diffusion-low indulgence characterizes the societies with aggressive deities to at least the .04 point of significance, employing the exact χ^2 solution.

In summary, there is a general tendency for less indulgent treatment in infancy to be related to predominantly aggressive deities in the cultural belief system, and for more indulgent treatment to be related to benevolent deities. The clearest relationship has to do with pain caused by the nurturing agents.

Childhood

The relations between the ratings of childhood variables and those of the supernaturals are examined in Table 4, which lists first "positive training" for the various systems of behavior, then punishments for failure to behave, then the reported actual frequency of children's behavior in the various systems. These relationships are followed by some compound indices of "pressure" (which combine the ratings of "positive training" and punishment for failure) and an index of "rigidity" of childhood training (the total score for punishment for nonperformance of all these behaviors).

The table shows that high self-reliance and independence training are related to the aggressiveness of the deities. This relationship holds for positive training, for punishment for nonperformance, for pressure, and for frequency of actual behavior. Despite an empirical relationship between the systems of self-reliance and independence training, we

TABLE 4
RELATIONSHIPS BETWEEN CHILDHOOD TRAINING VARIABLES AND THE
AGGRESSIVENESS AND BENEVOLENCE OF THE SUPERNATURALS

Childhood Training Variables	Supernaturals Predominantly		Significance Tests ^a
	Aggressive	Benevolent	
High Positive Training for Self-Reliance	62% (16/26) ^b	16% (3/19)	$p = <.05$; c.l. for 32/45 = .54-.83
High Positive Training for Independence	61% (17/28)	20% (4/20)	$p = <.05$; c.l. for 33/48 = .55-.82
High Positive Training for Nurturance	40% (8/20)	73% (11/15)	NS
High Positive Training for Responsibility	46% (12/26)	65% (13/20)	NS
High Positive Training for Obedience	33% (8/24)	44% (8/18)	NS
High Positive Training for Achievement	50% (11/22)	53% (8/15)	NS
High Punishment for Nonperf. Self-Reliance	73% (19/26)	31% (6/19)	$p = <.05$; c.l. for 32/45 = .54-.83
High Punishment for Nonperf. Independence	57% (16/28)	20% (4/20)	$p = <.05$; c.l. for 32/48 = .52-.80
High Punishment for Nonperf. Nurturance	60% (12/20)	71% (10/14)	NS
High Punishment for Nonperf. Responsibility	63% (15/24)	30% (6/20)	$p = <.10$; c.l. for 29/44 = .51-.80
High Punishment for Nonperf. Obedience	60% (15/25)	50% (9/18)	NS
High Punishment for Nonperf. Achievement	48% (10/21)	47% (7/15)	NS
High Frequency of Child's Self-Reliance	67% (18/27)	26% (5/19)	$p = <.05$; c.l. for 32/46 = .53-.82
High Frequency of Child's Independence	64% (18/28)	30% (6/20)	$p = <.05$; c.l. for 32/48 = .50-.59
High Frequency of Child's Nurturance	24% (5/21)	67% (10/15)	$p = <.02$; c.l. for 25/36 = .54-.85
High Frequency of Child's Responsibility	54% (13/24)	45% (9/20)	NS
High Frequency of Child's Obedience	44% (11/25)	56% (10/18)	NS
High Frequency of Achievement behavior	45% (10/22)	44% (7/16)	NS
High Pressure for Self-Reliance	73% (19/26)	32% (6/19)	$p = <.05$; c.l. for 32/45 = .54-.83
High Pressure for Independence	71% (20/28)	35% (7/20)	$p = <.05$; c.l. for 33/48 = .55-.82
High Pressure for Achievement	58% (11/19)	57% (8/14)	NS
High Rigidity Score	71% (12/17)	18% (2/11)	$p = <.01$; c.l. for 21/28 = .55-.89

^a We report significance level by χ^2 test, usually as .05 level with no further refinement. We also report 95% confidence limits for either the proportion of "successful" cases or of "unsuccessful" cases.

^b The figures in parentheses show the numbers on which the percentages are based. The numerator is the number of tribes showing the characteristic at the left out of the number of tribes classified as having either predominantly aggressive or benevolent gods (the denominator). All "high-low" breaks reported in this paper are the closest possible to the median.

have retained them as separate because of differences in meanings of the ratings.

The nurturance behavior system appears generally to be positively related to the benevolence of the supernaturals. This relationship is significant at beyond the .05 level for frequency of actual behavior, and the trend is maintained at levels short of statis-

tical significance for positive training and punishment for nonperformance.

There appears to be more positive training for responsibility behavior in societies with benevolent deities and more punishment for nonperformance in societies with aggressive deities, although neither of these relationships quite reaches statistical significance.

There is also a suggestive trend for the frequency of responsibility behavior to be somewhat higher in societies with aggressive deities. Thus responsibility training through reward may characterize societies with benevolent deities, and responsibility training through threat of punishment, those with aggressive deities. The obedience behavior system follows the same pattern as that for responsibility, except that the frequency of children's performance of obedience behavior tends (nonsignificantly) to be greater in societies with benevolent deities.

The achievement behavior system appears to be least related of any of the systems to these properties of the deities. In no case is there any discernible trend.

There is a tendency for societies with benevolent deities to be higher in positive training for four of the behavior systems, but with a very strong reversal on the remaining two—self-reliance and independence. The societies with aggressive deities, however, tend to use relatively more punishment for failure to perform in all but one of the systems—nurturance. This tendency in societies with aggressive deities toward control of behavior of children through punishment is most clearly highlighted in the rigidity score. Despite the general characterization of societies with aggressive deities as ones that make more use of punishment and less use of reward in socialization, it must be kept in mind that typical practices with respect to the various behavioral systems do differ. Societies with aggressive deities do reward self-reliance and independence (although this may be accompanied by some neglect), and societies with benevolent deities do tend to punish lapses from nurturance (although not significantly more than in the other kind of culture). To be ready to punish lapses does not necessarily betoken less attentiveness to children—it may require even more.

As far as children's actual performance of behavior in these systems is concerned, it appears to be clearly related to the properties of the supernaturals in three of the six systems. Children in societies with aggressive deities are more self-reliant, more independent, and less nurturant than those in societies with benevolent deities. The other directional

tendencies in children's behavior have already been noted.

DISCUSSION

The relationships that have been presented are useful in evaluating theories of the culture-personality relationship but are not decisive with regard to any of the major causal assumptions that such theory may take. One may, like McClelland (7), view the religious belief system as the independent factor, or one may see it as a projection of parental behavior. Another approach would view both the religious system and the child training system as controlled by some other aspect of the culture or personality of the people. And one may also see these relationships as specific historical accidents of no theoretical interest.

Our own interpretation of these data draws upon conditioning theory, reinforcement theory, and conflict theory (2). Let us trace in these terms the interrelations between the factual, psychological, and belief levels in societies with predominantly aggressive deities and those with predominantly benevolent deities. In societies with predominantly aggressive deities, we begin with the facts of hurt and pain in infancy, along with some nurture. On the psychological level, these facts should lead to anxiety in the child, because of his conflicting anticipations of hurt and of nurture. The resulting conflict, and attendant conflict drive, is reduced by a conception of the deity as more angry than kind and thus consonant with human anticipations of hurt. Concurrently with this resolution, we find on the psychological level a reduction in the tension of ambivalence in the child's anticipations and, in addition, a vicarious anxiety on the part of the parent for the child's welfare in a hurtful world. Returning to the factual level, we find the parent reinforcing independent and self-reliant behavior in his child to prepare him for the adult world, thus reducing the parent's own anxiety for the child.

Our interpretation of socialization in societies with predominantly benevolent deities starts in much the same way. We begin in infancy with the fact of considerable nurture, along with some hurt. The psychological con-

conflict between anticipations of nurture and of hurt is resolved here in the notion of a deity more kind than angry. Along with this belief we find, on the psychological level, a reduction in the tension arising from the ambivalence of the child's anticipations, and, in addition, vicarious anticipations in the parent of a probably pleasant future life for the child. On the factual level we find no pressures by the parent toward any particular behavior systems in the child but we do find considerable use of reward in child training, and we find the child, through identification or imitation, taking on such nurturant behavior.

An additional relationship that seems to require a different interpretation, though one not inconsistent with the foregoing, concerns the "capriciousness" of the gods. We assume that in societies with aggressive deities, the infant would not be able, particularly at the preverbal level, to understand or to predict the occasions of his receiving pain, and a property of capriciousness would thus accrue to the agent bestowing pain and nurture. A conception of the deities as capricious would therefore be seen as resolving the human anxiety deriving from inability to predict or foresee one's pains and woes. In another study, our raters were asked to judge whether a theme of "capriciousness" was present in the religious beliefs of a number of the same cultures as those analyzed in this study. A strong but nonsignificant tendency was found for the capricious deities to be the same as the aggressive ones, a relationship that becomes clearly significant when the variable of pain in infancy is controlled. Societies with predominantly aggressive deities and with high pain in the treatment of infants have capricious gods and spirits in six out of seven cases, and the societies with predominantly benevolent deities and with low pain in infant treatment lack capricious gods and spirits in seven out of seven cases ($p = <.004$).

We have considered two radically different interpretations of our data. The first of these would view the factual level of parents' training behavior toward infant and child as stemming from the religious belief system. That is, societies characterized by beliefs in predominantly aggressive deities would regulate their infant and child training practices

along compatibly aggressive lines. This view does not seem to us as fruitful of testable consequences as the interpretation proposed above, and there seems to be little evidence for or against it. We were able to make a minor check on it concerning the possibility that the pain involved in infant care would be explained or rationalized in the culture along explicitly religious lines. In none of the six societies with preponderantly aggressive deities for which data were easily available was the hurting of infants reported as done for religious purposes. This result is, of course, not conclusive, nor does it bear on the possibility of a "latent" religious theme as the psychological mediator of the practice of inflicting pain in infant care.

A second major interpretation would have its causal base in what we might call "nature." The aggressiveness of the deity, the amount of pain inflicted during infant care, and the emphasis or lack of emphasis in child training on particular behavior systems would all derive from the physical setting of the society. Unfortunately no satisfactory index is available that combines for each society the relevant aspects of climate, diet, energy and work levels, natural hazards, frequency of natural "calamities," etc. Lacking such an index that would permit a more definitive test, we used Horton's data (5) in a preliminary test of the hypothesis that both the aggressive properties of the supernaturals and the infancy and child training practices arise from the low subsistence level of a society. A small number of our societies also fall in Horton's sample, and his categorization of the level of subsistence "insecurity" is available for these cultures. There was no apparent relationship between subsistence insecurity and the properties of the deities. Another partial test employed some indices of aspects of natural phenomena in a number of our cultures provided by Whiting.⁶ One might entertain the possibility that belief in aggressive supernaturals is in part an outcome of extreme cold or of extreme heat conditions. No analyses that we have made to date show any clear relationships of this kind. The only trend is a weak directional one in which low mean temperatures (30° - 75°) go with aggressive supernaturals in seven out of ten cases, whereas

⁶ Whiting, J. W. M. Personal communication.

high mean temperatures (81° and up) tend to go with benevolent supernaturals in six out of eight cases. Temperature *variation* appears to have no trend of relationship with our data on the properties of the supernaturals.

Still additional explanatory hypotheses have been explored without appreciable support. According to one such hypothesis, mothers visit pain upon children as a displaced aggression arising from the frustrations of particularly low status. In a different study, our raters judged the status of women on such dimensions as ownership of property, inheritance of status, control of arrangements for love affairs and marriage, exercise of family authority, and so on. None of these, nor these in combination, predict the position of the societies on the variable of pain in infancy. The only suggestive trend is a tendency for the nurturant agent to inflict less pain on the infant in societies where the property is owned by women.

SUMMARY

The belief systems concerning supernatural beings of 62 societies with a wide geographic spread were characterized as being mainly aggressive or hurtful, or mainly benevolent. Other, often interrelated, factors in the socialization of the infant (approximately to a year and a half) or child (up to ten years) were related to the benevolence of the supernatural.

Societies with beliefs in aggressive supernaturals were significantly more likely than those with beliefs in benevolent gods and spirits to be described as having generally punitive or hurtful practices in treating infants. At levels short of statistical significance, such cultures had fewer nurturant agents, protected the infant less from environmental discomforts, showed him less affection, were more inconsistent in caring for his needs, and took less care of his needs. Societies with beliefs in aggressive supernaturals also tended to see their supernaturals as "capricious" in hurting people.

In regard to childhood, parents in societies with beliefs in aggressive supernaturals were found to be more likely to reward their children for self-reliance and independence and to

punish them for absence of these behaviors. They are generally more "rigid" in their training in the sense of depending more heavily on punishments than on rewards. Beliefs in benevolent gods and spirits are significantly related to the rated frequency of nurturant behavior shown by children in these cultures.

Several hypotheses according to which the obtained relationships might be explained were considered. The interpretation favored is derived from conditioning theory, reinforcement theory, and conflict theory. According to this view, the frequent hurt and pain in infancy in societies with aggressive deities causes anxiety in the child because of his conflicting anticipations of hurt and of nurture. His conflict is reduced by a conception of the deity as aggressive and thus compatible with human anticipations of hurt. The resulting reduction in the tension of ambivalence in the child's anticipations is accompanied by vicarious anxiety on the part of the parent concerning the child's future. The parent's anxiety is in turn reduced by following practices that reinforce independent and self-reliant behavior in his child to prepare him for the hurtful world he will encounter as an adult. In societies with benevolent deities there are no specific pressures toward training the child in particular behavior systems, but there is considerable reward used in child training. In result the child takes on such nurturant behavior through identification or imitation.

REFERENCES

1. BARRY, H., BACON, MARGARET K., & CHILD, I. L. A cross-cultural survey of some sex differences in socialization. *J. abnorm. soc. Psychol.*, 1957, **55**, 327-332.
2. BERLYNE, D. E. Uncertainty and conflict: A point of contact between information theory and behavior theory. *Psychol. Rev.*, 1957, **64**, 329-389.
3. CHILD, I. L., STORM, T., & VEROFF, J. An analysis of folk tales in relation to socialization practices. In J. W. Atkinson, *Motives in fantasy, action, and society: A method of assessment and study*. Princeton: D. Van Nostrand, 1958.
4. HALD, A., *Statistical tables and formulas*. New York: Wiley, 1952.
5. HEINICKE, C., & WHITING, B. B. *Bibliographies on personality and social development of the child*.

- New York: Social Science Research Council, Pamphlet 10, 1953.
6. HORTON, D. The functions of alcohol in primitive societies. *Quart. Jour. Stud. Alcohol*, 1943, 4, 292-303.
7. McCLELLAND, D. C. Some social consequences of achievement motivation. In M. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1955.
8. WHITING, J. W. M., CHILD, I. L., LAMBERT, W. W., et al. *Field guide for a study of socialization in five societies*. Laboratory of Human Development, 1955. (Mimeo.)

Received October 21, 1957.

ROLE, PERSONALITY, AND SOCIAL STRUCTURE IN THE ORGANIZATIONAL SETTING¹

DANIEL J. LEVINSON

Harvard University

DURING the past 20 years the concept of role has achieved wide currency in social psychology, sociology, and anthropology. From a sociopsychological point of view, one of its most alluring qualities is its double reference to the individual and to the collective matrix. The concept of role concerns the thoughts and actions of individuals, and, at the same time, it points up the influence upon the individual of socially patterned demands and standardizing forces. Partly for this reason, "role" has been seen by numerous writers (e.g., Gerth & Mills, 1953; Gross, Mason, & McEachern, 1958; Hartley & Hartley, 1952; Linton, 1945; Mead, 1934; Merton, 1957; Parsons, 1951; Sarbin, 1954) as a crucial concept for the linking of psychology, sociology, and anthropology. However, while the promise has seemed great, the fulfillment has thus far been relatively small. The concept of role remains one of the most overworked and underdeveloped in the social sciences.

My purpose here is to examine role theory primarily as it is used in the analysis of organizations (such as the hospital, business firm, prison, school). The organization provides a singularly useful arena for the development and application of role theory. It is small enough to be amenable to empirical study. Its structure is complex enough to provide a wide variety of social positions and role-standardizing forces. It offers an almost limitless opportunity to observe the individual personality *in vivo* (rather than in the psychologist's usual *vitro* of laboratory, survey questionnaire, or clinical office), selectively utilizing and modifying the demands and opportunities given in the social environment. The study of personality can, I submit, find no setting in which the

reciprocal impact of psyche and situation is more clearly or more dramatically evidenced.

Organizational theory and research has traditionally been the province of sociology and related disciplines that focus most directly upon the collective unit. Chief emphasis has accordingly been given to such aspects of the organization as formal and informal structure, administrative policy, allocation of resources, level of output, and the like. Little interest has been shown in the individual member as such or in the relevance of personality for organizational functioning. The prevailing image of the organization has been that of a mechanical apparatus operating impersonally once it is set in motion by administrative edict. The prevailing conception of social role is consonant with this image: the individual member is regarded as a cog in the apparatus, what he thinks and does being determined by requirements in the organizational structure.

This paper has the following aims: 1. To examine the traditional conception of organizational structure and role and to assess its limitations from a sociopsychological point of view. 2. To examine the conception of social role that derives from this approach to social structure and that tends, by definition, to exclude consideration of personality. 3. To provide a formulation of several, analytically distinct, role concepts to be used in place of the global term "role." 4. To suggest a theoretical approach to the analysis of relationships among role, personality, and social structure.

TRADITIONAL VIEWS OF BUREAUCRATIC STRUCTURE AND ROLE

Human personality has been virtually excluded from traditional organization theory. Its absence is perhaps most clearly reflected in Weber's (1946, 1947) theory of bureaucracy, which has become a major source of current thought regarding social organization and social role. I shall examine this theory briefly here, in order to point up some of its psychological limitations but without doing justice to its many virtues. In Weber's writings, the

¹ This is an expanded version of a paper presented at the meetings of the American Sociological Association in August, 1957. I am indebted to Peter D. Watson, Elizabeth Watson, and George Grosser for helpful criticism of the initial version. The ideas presented here stem in part from research supported by Grants M-687 and M-1000 from the National Institute of Mental Health, of the National Institutes of Health, Public Health Service.

bureaucratic organization is portrayed as a monolithic edifice. Norms are clearly defined and consistently applied, the agencies of role socialization succeed in inducing acceptance of organizational requirements, and the sanctions system provides the constraints and incentives needed to maintain behavioral conformity. Every individual is given a clearly defined role and readily "fills" it. There is little room in this tightly bound universe for more complex choice, for individual creativity, or for social change. As Gouldner (1954) has said of the studies carried out in this tradition: "Indeed, the social scene described has sometimes been so completely stripped of people that the impression is unintentionally rendered that there are disembodied social forces afoot, able to realize their ambitions apart from human action" (p. 16).

For Weber, bureaucracy as an ideal type is administered by "experts" in a spirit of impersonal rationality and is operated on a principle of discipline according to which each member performs his required duties as efficiently as possible. Rationality in decision-making and obedience in performance are the pivots on which the entire system operates. In this scheme of things, emotion is regarded merely as a hindrance to efficiency, as something to be excluded from the bureaucratic process.

The antipathy to emotion and motivation in Weber's thinking is reflected as well in his formulation of three types of authority: traditional, charismatic, and rational-legal. The rational-legal administrator is the pillar of bureaucracy. He receives his legitimation impersonally, from "the system," by virtue of his *technical* competence. His personal characteristics, his conception of the organization and its component groupings, his modes of relating to other persons (except that he be fair and impartial)—these and other psychological characteristics are not taken into theoretical consideration. There is no place in Weber's ideal type for the ties of affection, the competitive strivings, the subtle forms of support or of intimidation, so commonly found in even the most "rationalized" organizations. It is only the "charismatic" leader who becomes emotionally important to his followers and who must personally validate his right to lead.

While Weber has little to say about the

problem of motivation, two motives implicitly become universal instincts in his conception of "bureaucratic man." These are *conformity* (the motive for automatic acceptance of structural norms), and *status-seeking* (the desire to advance oneself by the acquisition and exercise of technical competence). More complex motivations and feelings are ignored.

There has been widespread acknowledgment of both the merits and the limitations of Weber's protean thought. However, the relevance of personality for organizational structure and role-definition remains a largely neglected problem in contemporary theory and research.² Our inadequacies are exemplified in the excellent *Reader in Bureaucracy*, edited by Merton, Gray, Hockey, and Selvin (1952). Although this book contains some of the most distinguished contributions to the field, it has almost nothing on the relation between organizational structure and personality. The editors suggest two lines of interrelation: first, that personality may be one determinant of occupational choice; and second, that a given type of structure may in time modify the personalities of its members. These are valuable hypotheses. However, they do not acknowledge the possibility that personality may have an impact on social structure. "The organization" is projected as an organism that either selects congenial personalities or makes over the recalcitrant ones to suit its own needs. This image is reflected in the editors' remark: "It would seem, therefore, that officials not initially suited to the demands of a bureaucratic position, progressively undergo modifications of personality" (p. 352). In other words, when social structure and personality fail to mesh, it

² Contemporary organization theory has benefited from criticisms and reformulations of Weber's theory by such writers as Barnard (1938), Friedrich (1950), Gerth and Mills (1953), Gouldner (1954), Merton (1957), and Parsons (in his introduction to Weber, 1947). Selznick (1957) has recently presented a conception of the administrative-managerial role that allows more room for psychological influences, but these are not explicitly conceptualized. There is growing though still inconclusive evidence from research on "culture and personality" work (Inkeles & Levinson, 1954) that social structures of various types both "require" and are influenced by modal personality, but this approach has received little application in research on organizations. An attempt at a distinctively sociopsychological approach, and a comprehensive view of the relevant literature, is presented by Argyris (1957).

is assumed to be personality alone that gives. Structure is the prime, uncaused, cause.

The impact of organizational structure on personality is indeed a significant problem for study. There is, however, a converse to this. When a member is critical of the organizational structure, he *may* maintain his personal values and traits, and work toward structural change. The manifold impact of personality on organizational structure and role remains to be investigated. To provide a theoretical basis for this type of investigation we need, I believe, to re-examine the concept of role.

"SOCIAL ROLE" AS A UNITARY CONCEPT

The concept of role is related to, and must be distinguished from, the concept of social position. A position is an element of organizational autonomy, a location in social space, a category of organizational membership. A role is, so to say, an aspect of organizational physiology; it involves function, adaptation, process. It is meaningful to say that a person "occupies" a social position; but it is inappropriate to say, as many do, that one occupies a role.

There are at least three specific senses in which the term "role" has been used, explicitly or implicitly, by different writers or by the same writer on different occasions.

a. Role may be defined as the *structurally given demands* (norms, expectations, taboos, responsibilities, and the like) associated with a given social position. Role is, in this sense, something outside the given individual, a set of pressures and facilitations that channel, guide, impede, support his functioning in the organization.

b. Role may be defined as the member's *orientation* or *conception* of the part he is to play in the organization. It is, so to say, his inner definition of what someone in his social position is supposed to think and do about it. Mead (1934) is probably the main source of this view of social role as an aspect of the person, and it is commonly used in analyses of occupational roles.

c. Role is commonly defined as the *actions* of the individual members—actions seen in terms of their relevance for the social structure (that is, seen in relation to the prevailing norms). In this sense, role refers to the ways in which members of a position act (with or without conscious intention) *in accord with or in violation of a given set of organizational norms*. Here, as in (b), role is defined as a characteristic of the actor rather than of his normative environment.

Many writers use a definition that embraces all of the above meanings without systematic distinction, and then shift, explicitly or im-

PLICITLY, from one meaning to another. The following are but a few of many possible examples.³

Each of the above three meanings of "role" is to be found in the writings of Parsons: (a) "From the point of view of the actor, his role is defined by the normative expectations of the members of the group as formulated in its social traditions" (Parsons, 1945, p. 230). (b) "The role is that organized sector of an actor's orientation which constitutes and defines his participation in an interactive process" (Parsons & Shils, 1951, p. 23). (c) "The status-role (is) the organized subsystem of acts of the actor or actors. . ." (Parsons, 1951, p. 26).

More often, the term is used in a way that includes all three meanings at once. In this *unitary*, all-embracing conception of role, there is, by assumption, a close fit between behavior and disposition (attitude, value), between societal prescription and individual adaptation. This point of view has its primary source in the writings of Linton, whose formulations of culture, status, and role have had enormous influence. According to Linton (1945), a role "includes the attitudes, values and behavior ascribed by the society to any and all persons occupying this status." In other words, society provides for each status or position a single mold that shapes the beliefs and actions of all its occupants.

Perhaps the most extensive formulation of this approach along sociopsychological lines is given by Newcomb (1950). Following Linton, Newcomb asserts, "Roles thus represent ways of carrying out the functions for which positions exist—ways which are generally agreed upon within (the) group" (p. 281). And, "Role is strictly a sociological concept; it purposely ignores individual, psychological facts" (p. 329). Having made this initial commitment to the "sociological" view that individual role-activity is a simple mirroring of group norms, Newcomb later attempts to find room for his "psychological" concerns with motivation,

³ An argument very similar to the one made here is presented by Gross, Mason, and McEachern (1958) in a comprehensive overview and critique of role theory. They point up the assumption of high consensus regarding role-demands and role-conceptions in traditional role theory, and present empirical evidence contradicting this assumption. Their analysis is, however, less concerned than the present one with the converging of role theory and personality theory.

meaning, and individual differences. He does this by partially giving up the "unitary" concept of role, and introducing a distinction between "prescribed role" and "role behavior." He avers that prescribed role is a sociological concept, "referring to common factors in the behaviors required" (p. 459), whereas role behavior is a psychological concept that refers to the activities of a single individual. The implications of this distinction for his earlier general definition of role are left unstated.

Whatever the merits or faults of Newcomb's reformulation, it at least gives conceptual recognition to the possibility that social prescription and individual adaptation may not match. This possibility is virtually excluded in the definition of social role forwarded by Linton and used by so many social scientists. In this respect, though certainly not in all respects, Linton's view is like Weber's: both see individual behavior as predominantly determined by the collective matrix. The matrix is, in the former case, culture, and in the latter, bureaucracy.

In short, the "unitary" conception of role assumes that there is a 1:1 relationship, or at least a *high degree of congruence*, among the three role aspects noted above. In the theory of bureaucratic organization, the rationale for this assumption is somewhat as follows. The organizationally given requirements will be internalized by the members and will thus be mirrored in their role-conceptions. People will know, and will want to do, what is expected of them. The agencies of role socialization will succeed except with a deviant minority—who constitute a separate problem for study. Individual action will in turn reflect the structural norms, since the appropriate role-conceptions will have been internalized and since the sanctions system rewards normative behavior and punishes deviant behavior. Thus, it is assumed that structural norms, individual role-conceptions and individual role-performance are three isomorphic reflections of a single entity: "the" role appropriate to a given organizational position.

It is, no doubt, reasonable to expect some degree of congruence among these aspects of a social role. Certainly, every organization contains numerous mechanisms designed to further such congruence. At the same time, it is a matter of common observation that organiza-

tions vary in the degree of their integration; structural demands are often contradictory, lines of authority may be defective, disagreements occur and reverberate at and below the surface of daily operations. To assume that what the organization requires, and what its members actually think and do, comprise a single, unified whole is severely to restrict our comprehension of organizational dynamics and change.

It is my thesis, then, that the unitary conception of social role is unrealistic and theoretically constricting. We should, I believe, eliminate the single term "role" except in the most general sense, i.e., of "role theory" as an over-all frame of analysis. Let us, rather, give independent conceptual and empirical status to the above three concepts and others. Let us investigate the relationships of each concept with the others, making no assumptions about the degree of congruence among them. Further, let us investigate their relationships with various other characteristics of the organization and of its individual members. I would suggest that the role concepts be named and defined as follows.

ORGANIZATIONALLY GIVEN ROLE-DEMANDS

The role-demands are external to the individual whose role is being examined. They are the situational pressures that confront him as the occupant of a given structural position. They have manifold sources: in the official charter and policies of the organization; in the traditions and ideology, explicit as well as implicit, that help to define the organization's purposes and modes of operation; in the views about this position which are held by members of the position (who influence any single member) and by members of the various positions impinging upon this one; and so on.

It is a common assumption that the structural requirements for any position are as a rule defined with a *high degree of explicitness, clarity, and consensus* among all the parties involved. To take the position of hospital nurse as an example: it is assumed that her role-requirements will be understood and agreed upon by the hospital administration, the nursing authorities, the physicians, etc. Yet one of the striking research findings in all manner of hospitals is the failure of consensus

regarding the proper role of nurse (e.g., Burling, Lentz, & Wilson, 1956; Argyris, 1957). Similar findings have been obtained in school systems, business firms, and the like (e.g., Gross et al., 1958; Kornhauser, Dubin, & Ross, 1954).

In attempting to characterize the role-requirements for a given position, one must therefore guard against the assumption that they are unified and logically coherent. There may be major differences and even contradictions between official norms, as defined by charter or by administrative authority, and the "informal" norms held by various groupings within the organization. Moreover, within a given-status group, such as the top administrators, there may be several conflicting viewpoints concerning long range goals, current policies, and specific role-requirements. In short, the structural demands themselves are often multiple and disunified. Few are the attempts to investigate the sources of such disunity, to acknowledge its frequency, or to take it into conceptual account in general structural theory.

It is important also to consider the specificity or *narrowness* with which the normative requirements are defined. Norms have an "ought" quality; they confer legitimacy and reward-value upon certain modes of action, thought and emotion, while condemning others. But there are degrees here. Normative evaluations cover a spectrum from "strongly required," through various degrees of qualitative kinds of "acceptable," to more or less stringently tabooed. Organizations differ in the width of the intermediate range on this spectrum. That is, they differ in the number and kinds of adaptation that are normatively acceptable. The wider this range—the less specific the norms—the greater is the area of personal choice for the individual. While the existence of such an intermediate range is generally acknowledged, structural analyses often proceed as though practically all norms were absolute prescriptions or proscriptions allowing few alternatives for individual action.

There are various other normative complexities to be reckoned with. A single set of role-norms may be internally contradictory. In the case of the mental hospital nurse, for example, the norm of maintaining an "orderly ward" often conflicts with the norm of en-

couraging self-expression in patients. The individual nurse then has a range of choice, which may be narrow or wide, in balancing these conflicting requirements. There are also ambiguities in norms, and discrepancies between those held explicitly and those that are less verbalized and perhaps less conscious. These normative complexities permit, and may even induce, significant variations in individual role-performance.

The degree of *coherence* among the structurally defined role-requirements, the degree of *consensus* with which they are held, and the degree of *individual choice* they allow (the range of acceptable alternatives) are among the most significant properties of any organization. In some organizations, there is very great coherence of role-requirements and a minimum of individual choice. In most cases, however, the degree of integration within roles and among sets of roles appears to be more moderate.⁴ This structural pattern is of especial interest from a sociopsychological point of view. To the extent that the requirements for a given position are ambiguous, contradictory, or otherwise "open," the individual members have greater opportunity for selection among existing norms and for creation of new norms. In this process, personality plays an important part. I shall return to this issue shortly.

While the normative requirements (assigned tasks, rules governing authority-subordinate relationships, demands for work output, and the like) are of great importance, there are other aspects of the organization that have an impact on the individual member. I shall mention two that are sometimes neglected.

Role-facilities. In addition to the demands and obligations imposed upon the individual, we must also take into account the techniques, resources, and conditions of work—the means made available to him for fulfilling his organizational functions. The introduction of tranquillizing drugs in the mental hospital, or of automation in industry, has provided tre-

⁴ The reduced integration reflects in part the tremendous rate of technological change, the geographical and occupational mobility, and the diversity in personality that characterize modern society. On the other hand, diversity is opposed by the standardization of culture on a mass basis and by the growth of large-scale organization itself. Trends toward increased standardization and uniformity are highlighted in Whyte's (1956) analysis.

mendous leverage for change in organizational structure and role-definition. The teacher-student ratio, an ecological characteristic of every school, grossly affects the probability that a given teacher will work creatively with individual students. In other words, technological and ecological facilities are not merely "tools" by which norms are met; they are often a crucial basis for the maintenance or change of an organizational form.

Role-dilemmas or problematic issues. In describing the tasks and rules governing a given organizational position, and the facilities provided for their realization, we are, as it were, looking at that position from the viewpoint of a higher administrative authority whose chief concern is "getting the job done." Bureaucracy is often analyzed from this (usually implicit) viewpoint. What is equally necessary, though less often done, is to look at the situation of the position-members from their own point of view: the meaning it has for them, the feelings it evokes, the ways in which it is stressful or supporting. From this sociopsychological perspective, new dimensions of role analysis emerge. The concept of role-dilemma is an example. The usefulness of this concept stems from the fact that every human situation has its contradictions and its problematic features. Where such dilemmas exist, there is no "optimal" mode of adaptation; each mode has its advantages and its costs. Parsons (1951), in his discussion of "the situation of the patient," explores some of the dilemmas confronting the ill person in our society. Erikson (1957) and Pine and Levinson (1958) have written about the dilemmas of the mental hospital patient; for example, the conflicting pressures (from without and from within) toward cure through self-awareness and toward cure through repressive self-control. Role-dilemmas of the psychiatric resident have been studied by Sharaf and Levinson (1957). Various studies have described the problems of the factory foreman caught in the conflicting cross-pressures between the workers he must supervise and the managers to whom he is responsible. The foreman's situation tends to evoke feelings of social marginality, mixed identifications, and conflicting tendencies to be a good "older brother" with subordinates and an obedient son with higher authority. Role-dilemmas have their sources both in

organizational structure and in individual personality. Similarly, both structure and personality influence the varied forms of adaptation that are achieved. The point to be emphasized here is that every social structure confronts its members with adaptive dilemmas. If we are to comprehend this aspect of organizational life, we must conceive of social structure as having intrinsically *psychological* properties, as making complex psychological demands that affect, and are affected by, the personalities of its members.

PERSONAL ROLE-DEFINITION

In the foregoing we have considered the patterning of the environment for an organizational position—the kind of sociopsychological world with which members of the position must deal. Let us turn now to the individual members themselves. Confronted with a complex system of requirements, facilities, and conditions of work, the individual effects his modes of adaptation. I shall use the term "personal role-definition" to encompass the individual's adaptation within the organization. This may involve passive "adjustment," active furthering of current role-demands, apparent conformity combined with indirect "sabotage," attempts at constructive innovation (revision of own role or of broader structural arrangements), and the like. The personal role-definition may thus have varying degrees of fit with the role-requirements. It may serve in various ways to maintain or to change the social structure. It may involve a high or a low degree of self-commitment and personal involvement on the part of the individual (Selznick, 1957).

For certain purposes, it is helpful to make a sharp distinction between two levels of adaptation: at a more *ideational* level, we may speak of a role-conception; at a more *behavioral* level, there is a pattern of role-performance. Each of these has an affective component. Role-conception and role-performance are independent though related variables; let us consider them in turn.

Individual (and modal) role-conceptions. The nature of a role-conception may perhaps be clarified by placing it in relation to an ideology. The boundary between the two is certainly not a sharp one. However, ideology refers most directly to an orientation regarding the entire

organizational (or other) structure—its purposes, its modes of operation, the prevailing forms of individual and group relationships, and so on. A role-conception offers a definition and rationale for one position within the structure. If ideology portrays and rationalizes the organizational world, then role-conception delineates the specific functions, values, and manner of functioning appropriate to one position within it.

The degree of uniformity or variability in individual role-conceptions within a given position will presumably vary from one organization to another. When one or more types of role-conception are commonly held (consensual), we may speak of modal types. The maintenance of structural stability requires that there be at least moderate consensus and that modal role-conceptions be reasonably congruent with role-requirements. At the same time, the presence of incongruent modal role-conceptions may, under certain conditions, provide an ideational basis for major organizational change.

Starting with the primary assumption that each member "takes over" a structurally defined role, many social scientists tend to assume that there is great uniformity in role-conception among the members of a given social position. They hold, in other words, that for every position there is a *dominant, modal role-conception corresponding to the structural demands*, and that there is relatively little individual deviation from the modal pattern. Although this state of affairs may at times obtain, we know that the members of a given social position often have quite diverse conceptions of their proper roles (Greenblatt, Levinson, & Williams, 1957; Gross, Mason, & McEachern, 1958; Reissman & Rohrer, 1957; Bendix, 1956). After all, individual role-conceptions are formed only partially within the present organizational setting. The individual's ideas about his occupational role are influenced by childhood experiences, by his values and other personality characteristics, by formal education and apprenticeship, and the like. The ideas of various potential reference groups within and outside of the organization are available through reading, informal contacts, etc. There is reason to expect, then, that the role-conceptions of individuals in a given organizational position will vary and

will not always conform to official role-requirements. Both the diversities and the modal patterns must be considered in organizational analysis.

Individual (and modal) role-performance. This term refers to the overt behavioral aspect of role-definition—to the more or less characteristic ways in which the individual acts as the occupant of a social position. Because role-performance involves immediately observable behavior, its description would seem to present few systematic problems. However, the formulation of adequate variables for the analysis of role-performance is in fact a major theoretical problem and one of the great stumbling blocks in empirical research.

Everyone would agree, I suppose, that role-performance concerns only those aspects of the total stream of behavior that are structurally relevant. But which aspects of behavior are the important ones? And where shall the boundary be drawn between that which is structurally relevant and that which is incidental or idiosyncratic?

One's answer to these questions probably depends, above all, upon his conception of social structure. Those who conceive of social structure rather narrowly in terms of concrete work tasks and normative requirements, are inclined to take a similarly narrow view of role. In this view, role-performance is simply the fulfillment of formal role-norms, and anything else the person does is extraneous to role-performance as such. Its proponents acknowledge that there are variations in "style" of performance but regard these as incidental. What is essential to *role-performance* is the degree to which norms are met.

A more complex and inclusive conception of social structure requires correspondingly multi-dimensional delineation of role-performance. An organization has, from this viewpoint, "latent" as well as "manifest" structure; it has a many-faceted emotional climate; it tends to "demand" varied forms of interpersonal allegiance, friendship, deference, intimidation, ingratiation, rivalry, and the like. If characteristics such as these are considered intrinsic properties of social structure, then they must be included in the characterization of role-performance. My own preference is for the more inclusive view. I regard social structure as having psychological as well as

other properties, and I regard as intrinsic to role-performance the varied meanings and feelings which the actor communicates to those about him. Ultimately, we must learn to characterize organizational behavior in a way that takes into account, and helps to illuminate, its functions for the individual, for the others with whom he interacts, and for the organization.

It is commonly assumed that there is great uniformity in role-performance among the members of a given position. Or, in other words, that there is a *dominant, modal pattern of role-performance corresponding to the structural requirements*. The rationale here parallels that given above for role-conceptions. However, where individual variations in patterns of role-performance have been investigated, several modal types rather than a single dominant pattern were found (Argyris, 1957; Greenblatt et al., 1957).

Nor is this variability surprising, except to those who have the most simplistic conception of social life. Role-performance, like any form of human behavior, is the resultant of many forces. Some of these forces derive from the organizational matrix; for example, from role-demands and the pressures of authority, from informal group influences, and from impending sanctions. Other determinants lie within the person, as for example his role-conceptions and role-relevant personality characteristics. Except in unusual cases where all forces operate to channel behavior in the same direction, role-performance will reflect the individual's attempts at choice and compromise among diverse external and internal forces.

The relative contributions of various forms of influence to individual or modal role-performance can be determined only *if each set of variables is defined and measured independently of the others*. That is, indeed, one of the major reasons for emphasizing and sharpening the distinctions among role-performance, role-conception, and role-demands. Where these distinctions are not sharply drawn, there is a tendency to study one element and to assume that the others are in close fit. For example, one may learn from the official charter and the administrative authorities how the organization is supposed to work—the formal requirements—and then assume that it in fact operates in this way. Or, conversely, one may

observe various regularities in role-performance and then assume that these are structurally determined, without independently assessing the structural requirements. To do this is to make structural explanations purely tautologous.

More careful distinction among these aspects of social structure and role will also, I believe, permit greater use of personality theory in organizational analysis. Let us turn briefly to this question.

ROLE-DEFINITION, PERSONALITY, AND SOCIAL STRUCTURE

Just as social structure presents massive forces which influence the individual from without toward certain forms of adaptation, so does personality present massive forces from within which lead him to select, create, and synthesize certain forms of adaptation rather than others. Role-definition may be seen from one perspective as an aspect of personality. It represents the individual's attempt to structure his social reality, to define his place within it, and to guide his search for meaning and gratification. Role-definition is, in this sense, an *ego achievement*—a reflection of the person's capacity to resolve conflicting demands, to utilize existing opportunities and create new ones, to find some balance between stability and change, conformity and autonomy, the ideal and the feasible, in a complex environment.

The formation of a role-definition is, from a dynamic psychological point of view, an "external function" of the ego. Like the other external (reality-oriented) ego functions, it is influenced by the ways in which the ego carries out its "internal functions" of coping with, and attempting to synthesize, the demands of id, superego, and ego. These internal activities—the "psychodynamics" of personality—include among other things: unconscious fantasies; unconscious moral conceptions and the wishes against which they are directed; the characteristic ways in which unconscious processes are transformed or deflected in more conscious thought, feeling, and behavioral striving; conceptions of self and ways of maintaining or changing these conceptions in the face of changing pressures from within and from the external world.

In viewing role-definition as an aspect of

personality, I am suggesting that it is, to *varying degrees*, related to and imbedded within other aspects of personality. An individual's conception of his role in a particular organization is to be seen within a series of wider psychological contexts: his conception of his occupational role generally (occupational identity), his basic values, life-goals, and conception of self (ego identity), and so on. Thus, one's way of relating to authorities in the organization depends in part upon his relation to authority in general, and upon his fantasies, conscious as well as unconscious, about the "good" and the "bad" parental authority. His ways of dealing with the stressful aspects of organizational life are influenced by the impulses, anxieties, and modes of defense that these stresses activate in him (Argyris, 1957; Erikson, 1950; Henry, 1949; Blum, 1933; Pine & Levinson, 1957).

There are variations in the degree to which personal role-definition is imbedded in, and influenced by, deeper-lying personality characteristics. The importance of individual or modal personality for role-definition is a matter for empirical study and cannot be settled by casual assumption. Traditional sociological theory can be criticized for assuming that individual role-definition is determined almost entirely by social structure. Similarly, dynamic personality theory will not take its rightful place as a crucial element of social psychology until it views the individual within his sociocultural environment. Lacking an adequate recognition and *conceptualization* of the individual's external reality—including the "reality" of social structure—personality researchers tend to assume that individual adaptation is primarily personality-determined and that reality is, for the most part, an amorphous blob structured by the individual to suit his inner needs.

Clearly, individual role-conception and role-performance do not emanate, fully formed, from the depths of personality. Nor are they simply mirror images of a mold established by social structure. Elsewhere (Levinson, 1954), I have used the term "mirage" theory for the view, frequently held or implied in the psychoanalytic literature, that ideologies, role-conceptions, and behavior are mere epiphenomena or by-products of unconscious fantasies and defenses. Similarly, the term "sponge" theory

characterizes the view, commonly forwarded in the sociological literature, in which man is merely a passive, mechanical absorber of the prevailing structural demands.

Our understanding of personal role-definition will remain seriously impaired as long as we fail to place it, analytically, in *both intra-personal and structural-environmental contexts*. That is to say, we must be concerned with the meaning of role-definition both for the individual personality and for the social system. A given role-definition is influenced by, and has an influence upon, the *psyche* as well as the *socius*. If we are adequately to understand the nature, the determinants, and the consequences of role-definition, we need the double perspective of personality and social structure. The use of these two reference points is, like the use of our two eyes in seeing, necessary for the achievement of depth in our social vision.

Theory and research on organizational roles must consider relationships among at least the following sets of characteristics: structurally given role-demands and -opportunities, personal role-definition (including conceptions and performance), and personality in its role-related aspects. Many forms of relationship may exist among them. I shall mention only a few hypothetical possibilities.

In one type case, the role-requirements are so narrowly defined, and the mechanisms of social control so powerful, that only one form of role-performance can be sustained for any given position. An organization of this type may be able selectively to recruit and retain only individuals who, by virtue of personality, find this system meaningful and gratifying. If a congruent modal personality is achieved, a highly integrated and stable structure may well emerge. I would hypothesize that a structurally congruent modal personality is one condition, though by no means the only one, for the stability of a rigidly integrated system. (In modern times, of course, the rapidity of technological change prevents long-term stability in any organizational structure.)

However, an organization of this kind may acquire members who are not initially receptive to the structural order, that is, who are *incongruent* in role-conception or in personality. Here, several alternative developments are possible.

1. The "incongruent" members may change so that their role-conceptions and personalities come better to fit the structural requirements.

2. The incongruent ones may leave the organization, by choice or by expulsion. The high turnover in most of our organizations is due less to technical incompetence than to rejection of the "conditions of life" in the organization.

3. The incongruent ones may remain, but in a state of apathetic conformity. In this case, the person meets at least the minimal requirements of role-performance but his role-conceptions continue relatively unchanged, he gets little satisfaction from work, and he engages in repeated "sabotage" of organizational aims. This is an uncomfortably frequent occurrence in our society. In the Soviet Union as well, even after 40 years of enveloping social controls, there exist structurally incongruent forms of political ideology, occupational role-definition, and personality (Inkeles, Hanfmann, & Beier, 1958).

4. The incongruent members may gain sufficient social power to change the organizational structure. This phenomenon is well known, though not well enough understood. For example, in certain of our mental hospitals, schools and prisons over the past 20-30 years, individuals with new ideas and personal characteristics have entered in large enough numbers, and in sufficiently strategic positions, to effect major structural changes. Similar ideological and structural transitions are evident in other types of organization, such as corporate business.

The foregoing are a few of many possible developments in a relatively monolithic structure. A somewhat looser organizational pattern is perhaps more commonly found. In this setting, structural change is seen as a legitimate function of members at various levels in the organization. To the extent that diversity and innovation are valued (rather than merely given lip-service), variations in individual role-definition are tolerated or even encouraged within relatively wide limits. The role-definitions that develop will reflect various degrees of synthesis and compromise between personal preference and structural demand.

In summary, I have suggested that a primary distinction be made between the structurally given role-demands and the forms of role-definition achieved by the individual members of an organization. Personal role-definition then becomes a linking concept between personality and social structure. It can be seen as a reflection of those aspects of individual personality that are activated and sustained in a given structural-ecological environment. This view is opposed both to the "sociologizing" of individual behavior and to

the "psychologizing" of organizational structure. At the same time, it is concerned with both the psychological properties of social structure and the structural properties of individual adaptation.

Finally, we should keep in mind that both personality structure and social structure inevitably have their internal contradictions. No individual is sufficiently all of a piece that he will for long find any form of adaptation, occupational or otherwise, totally satisfying. Whatever the psychic gains stemming from a particular role-definition and social structure, there will also be losses: wishes that must be renounced or made unconscious, values that must be compromised, anxieties to be handled, personal goals that will at best be incompletely met. The organization has equivalent limitations. Its multiple purposes cannot all be optimally achieved. It faces recurrent dilemmas over conflicting requirements: control and freedom; centralization and decentralization of authority; security as against the risk of failure; specialization and diffusion of work function; stability and change; collective unity and diversity. Dilemmas such as these arise anew in different forms at each new step of organizational development, without permanent solution. And perpetual changes in technology, in scientific understanding, in material resources, in the demands and capacities of its members and the surrounding community, present new issues and require continuing organizational readjustment.

In short, every individual and every sociocultural form contains within itself the seeds of its own destruction—or its own reconstruction. To grasp both the sources of stability and the seeds of change in human affairs is one of the great challenges to contemporary social science.

REFERENCES

- ARGYRIS, C. *Human relations in a hospital*. New Haven: Labor and Management Center, 1955.
- ARGYRIS, C. *Personality and organization*. New York: Harper, 1957.
- BARNARD, C. I. *The functions of the executive*. Cambridge, Mass.: Harvard Univer. Press, 1938.
- BENDIX, R. *Work and authority in industry*. New York: Wiley, 1956.
- BLUM, F. H. *Toward a democratic work process*. New York: Harper, 1933.
- BURLING, T., LENTZ, EDITH, & WILSON, R. N. *The give and take in hospitals*. New York: Putnam, 1956.

- ERIKSON, E. H. *Childhood and society*. New York: Norton, 1950.
- ERIKSON, K. T. Patient role and social uncertainty: A dilemma of the mentally ill. *Psychiatry*, 1957, 20, 263-274.
- FRIEDRICH, C. J. *Constitutional government and democracy*. Boston: Little Brown, 1950.
- GERTH, H. H. & MILLS, C. W. *Character and social structure*. New York: Harcourt, Brace, 1953.
- GOULDNER, A. W. *Patterns of industrial bureaucracy*. Glencoe, Ill.: Free Press, 1954.
- GREENBLATT, M., LEVINSON, D. J., & WILLIAMS, R. H. (Eds.) *The patient and the mental hospital*. Glencoe, Ill.: Free Press, 1957.
- GROSS, N., MASON, W. S., & MCEACHERN, A. W. *Explorations in role analysis*. New York: Wiley, 1958.
- HARTLEY, E. L., & HARTLEY, RUTH E. *Fundamentals of social psychology*. New York: Knopf, 1952.
- HENRY, W. E. The business executive: the psychodynamics of a social role. *Amer. J. Sociol.*, 1949, 54, 286-91.
- INKELES, A., HANFMANN, EUGENIA, & BEIER, HELEN. Modal personality and adjustment to the Soviet political system. *Hum. Relat.*, 1958, 11, 3-22.
- INKELES, A., & LEVINSON, D. J. National character: The study of modal personality and socio-cultural systems. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954.
- KORNHAUSER, A., DUBIN, R. & ROSS, A. M. *Industrial conflict*. New York: McGraw Hill, 1954.
- LEVINSON, D. J. *Idea systems in the individual and society*. Paper presented at Boston University, Founder's Day Institute, 1954. Mimeographed: Center for Sociopsychological Research, Massachusetts Mental Health Center.
- LINTON, R. *The cultural background of personality*. New York: Appleton-Century, 1945.
- MEAD, G. H. *Mind, self and society*. Chicago: Univer. Chicago Press, 1934.
- MERTON, R. K. *Social theory and social structure*. (Rev. Ed.) Glencoe, Ill.: Free Press, 1957.
- MERTON, R. K., GRAY, A. P., HOCKEY, BARBARA, & SELVIN, H. C. *Reader in bureaucracy*. Glencoe, Ill.: Free Press, 1957.
- NEWCOMB, T. M. *Social psychology*. New York: Dryden, 1950.
- PARSONS, T. *Essays in sociological theory*. (Rev. ed.) Glencoe, Ill.: Free Press, 1945.
- PARSONS, T. *The social system*. Glencoe, Ill.: Free Press, 1951.
- PARSONS, T., & SHILS, E. A. (Eds.) *Toward a general theory of action*. Cambridge, Mass.: Harvard Univer. Press, 1951.
- PINE, F., & LEVINSON, D. J. Two patterns of ideology, role conception, and personality among mental hospital aides. In M. Greenblatt, D. J. Levinson, & R. H. Williams (Eds.), *The patient and the mental hospital*. Glencoe, Ill.: Free Press, 1957.
- PINE, F., & LEVINSON, D. J. *Problematic issues in the role of mental hospital patient*. Mimeographed: Center for Sociopsychological Research, Massachusetts Mental Health Center, 1958.
- REISSMAN, L., & ROHRER, J. J. (Eds.) *Change and dilemma in the nursing profession*. New York: Putnam, 1957.
- SARBIN, T. R. Role theory. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954.
- SELZNICK, P. *Leadership in administration*. Evanston, Ill.: Row, Peterson, 1957.
- SHARAF, M. R., & LEVINSON, D. J. Patterns of ideology and role definition among psychiatric residents. In M. Greenblatt, D. J. Levinson, & R. H. Williams, (Eds.), *The patient and the mental hospital*. Glencoe, Ill.: Free Press, 1957.
- WEBER, M. *Essays in sociology*. Ed. by H. H. Gerth & C. W. Mills. New York: Oxford Univer. Press, 1946.
- WEBER, M. *The theory of social and economic organization*. Ed. by T. Parsons. New York: Oxford Univer. Press, 1947.
- WHYTE, W. F. *The organization man*. New York: Simon & Shuster, 1956.

Received November 27, 1957.

TEST-ANXIETY AND BLAME-ASSIGNMENT IN GRADE SCHOOL CHILDREN¹

JOHN DORIS

Yale University

DESPITE the increasing use of self-rated scales of anxiety in the study of learning and other intellectual performances (Child, 1954; Taylor, 1956), relatively little study has been made of the relationship of these self-rated scales to attitudinal reactions or sets of the subject (S) toward himself and others in regard to these performances. Yet the study of such attitudinal reactions and sets must be considered basic to an ultimate understanding of the way in which anxiety-scale scores relate to performance. It seems self-evident that the complex of attitudes which surrounds his performance must of necessity interact with S's anxiety concerning his performance and the efficiency of the performance itself.

Previous research and theoretical considerations (Doris, 1957; Doris & Sarason, 1955; Sarason & Mandler, 1952; Dollard, Miller, Mowrer, Sears, Ford, Hovland, & Sollenberger, 1939; Freud, 1946), lead us to expect that those attitudes which may be classified as hostile or aggressive are especially relevant to the relationship of anxiety and performance. In a study utilizing college Ss, Doris and Sarason (1955) attempted to demonstrate a relationship between test-anxiety and blame-assignment in a failure situation. The Ss were individually administered various tasks with success and failure predetermined by the experimenter (E). Differences in self-blame assignment between high-test-anxious and low-test-anxious Ss were noted under one of the conditions of test administration but not under a second. It was hypothesized that the major difference between the two conditions of administration was in the amount of stress imposed on the Ss.

¹ This paper is adapted from a dissertation submitted to Yale University in partial fulfillment of the requirements for the Ph.D. degree. The author wishes to express his deep appreciation to Seymour Sarason, Robert Abelson, and the late Katherine Wolf for their many constructive criticisms in the design and execution of this investigation. Special thanks are due to Robert Abelson for his advice on the statistical analysis.

The subsequent development of a Childrens' Text-Anxiety (CTA) scale² (Sarason, Davidson, Lighthall, & Waite, in press) offered the opportunity for further study of this problem of test-anxiety and blame-assignment. In extending the investigation from a college population to elementary school children, attention was focused on the hypothesis that in a failure situation high-test-anxiety Ss tend to assign blame for failure to themselves to a greater extent than low-test-anxious Ss. An attempt to answer this basic question could be undertaken by a simple repetition of the college study, using elementary school Ss. However, it was decided that a somewhat different approach might result in greater experimental economy by throwing light upon related questions at the same time it sought an answer to the basic one.

Consequently, an investigation consisting of four studies was undertaken with fifth- and sixth-grade children in the Milford, Connecticut, public school system.³ The first three of these studies were considered pretests for the fourth. The results of these pretests were such as to place the problem of blame-assignment and test-anxiety in a different perspective from that in which it was viewed at the start of the investigation. Therefore, these pretests will be briefly summarized.

PRETESTS

Study I. Six groups, each consisting of either eight boys or eight girls, were presented with an arithmetic multiplication test. The first child in his group who finished all problems correctly was declared the winner. The winner was then removed from the group, and the remaining Ss were asked to fill out a question-

² The CTA scale is a questionnaire of 42 items relating to children's reactions to tests and classroom recitations. The scale is presented orally and the children indicate their responses by circling a Yes or No for each item.

³ The author wishes to express his thanks to Irving Zweibelson and the teachers of the Milford Public School System for their cooperation with this study.

naire which gave them opportunity to assign blame for their poor performance to themselves or to others. Each of the 16 items on this questionnaire was answered with a yes or no response. Typical items were as follows:

Yes No 6. I don't pay enough attention when arithmetic is being taught.

Yes No 9. I might have won if we had been told more clearly what we should do before we began to work.

After collection of the questionnaire, another arithmetic test was given and another winner selected. The questionnaire was then readministered. Finally, a third arithmetic test and a third administration of the questionnaire were undertaken.

Each group now consisted of five Ss who had undergone three successive failures in a competitive arithmetic test and who had filled out a blame questionnaire after each failure. Of 30 such Ss, CTA scores were available for 22.⁴ Self-blame (SB) and other-blame (OB) scores were obtained by adding the number of yes responses to self-blame and other-blame items for all three presentations of the questionnaire. Blame scores and CTA scores were then correlated.⁵ The obtained correlations were as follows: CTA-SB, .596 ($p < .01$); CTA-OB, .481 ($p < .05$); SB-OB, .472 ($p < .05$).

In view of the unexpected correlations between CTA and OB and SB and OB, the question arose as to the possible presence of an artifact in the use of the questionnaires, such as a generalized response tendency to questionnaires of a similar yes and no format.

Study II. The format of the blame scales was revised. For each item of self-blame or other-blame there was a corresponding item of non-blame. These 32 blame and nonblame items

were presented in pairs. Examples are as follows:

—3a. I paid strict attention to the instructions.
—3b. I didn't pay enough attention to the instructions.

—7a. I think the test was fair.
—7b. I think the test was not fair.

As *E* read each pair of items, the children were instructed to mark an *X* beside that member of a pair that most nearly expressed the way that they felt. Again, groups of eight were given arithmetic problems and three winners selected from each group, leaving five who had experienced three successive failures and who had marked three blame questionnaires. In addition, Ss were given a questionnaire which gave them the opportunity to express dissatisfaction or aggression toward themselves or others in relationship to school situations or activities. Two examples follow:

Yes No 1. Even if I try I can't write as neatly as other children.

Yes No 2. The teacher should speak more loudly when reading to the class.

The purpose of this scale was to determine if there are more generalized attitudes of dissatisfaction and aggression in relation to school which correlate with CTA scores. CTA scores were available for 23 children of those who had experienced three consecutive failures. The results of this pretest gave some encouragement for the belief that the relationships between CTA and SB and CTA and OB were not an artifact of the scale format. Despite a radical change in the format of the blame scales, the relationship between CTA and SB (.425, $p < .05$) and CTA and OB (.530, $p < .01$) remained statistically significant. However the correlation between SB and OB (.116) was not significant.

From the school-dissatisfaction scale, scores of self-dissatisfaction and other-dissatisfaction in relationship to school were obtained. It was determined that CTA scores correlated with both self-dissatisfaction (.546, $p < .01$) and other-dissatisfaction (.645, $p < .01$). This raised the question as to whether or not the self-blame and other-blame attitudes in failure are only a reflection of more general attitudes of dissatisfaction or aggression. Allied to this

⁴ Test-anxiety scores for the Ss in the present investigation were available from questionnaires administered in the spring of 1955. It was decided to make use of these scores as our measure of test-anxiety, despite a lapse of a year's time, rather than to administer anew the CTA questionnaire, because considerations of efficiency in the over-all CTA project made it undesirable to administer the scale separately for each subproject and because it seemed desirable to have the present investigation of blame-assignment distinct in S's mind from his experience with the CTA scale.

⁵ All correlation coefficients reported in this investigation are product-moment correlation coefficients.

was the question of whether or not failure has any experimental effect on the relationship of CTA and the assignment of blame in a failure situation.

Study III. In the third study, the blame scale was changed so that items of blame did not necessarily refer to a specific test experience. Examples are as follows:

- 3a. When I take an arithmetic test I pay strict attention to the instructions.
- 3b. Sometimes when I take an arithmetic test I don't pay enough attention to the instructions.
- 7a. I think that arithmetic tests are always fair.
- 7b. I think that arithmetic tests are sometimes unfair.

The scale included a total of 11 self-blame and 11 other-blame items, each paired with a corresponding nonblame item. Again six groups of eight were given competitive arithmetic tests in which the first *S* finished correctly was declared the winner. Both before and after the test the blame scale was administered to the *Ss*. Therefore, in each group there were seven *Ss* who had filled out the blame questionnaire both before and after a specific failure in arithmetic. Of these 42 children, CTA scores were available for 36. From the blame questionnaire it was possible to obtain before-failure-self-blame, after-failure-self-blame, before-failure-other-blame, and after-failure-other-blame scores. These scores were correlated with CTA scores. The correlations of CTA with post-failure-self-blame (.234) and post-failure-other-blame (.317) were obviously not greater than pre-failure-self-blame (.437, $p < .05$) and pre-failure-other-blame (.499, $p < .01$). Therefore, the experimental variable of failure could not be affecting the relationship of CTA and blame-assignment in the expected positive direction.

THE PROBLEM

The findings of Studies I, II, and III posed two problems. First, since the relationship between text-anxiety and blame scores was not experimentally manipulated in the case of the school children, is it possible that this relationship is an artifact of the questionnaire technique, e.g., some general response tendencies of *Ss* to questionnaires of this type? This, of

course, is the question which arose with Study I and to which Study II attempted to give some answer by changing the format of the blame scale. But with the failure to control the relationship in Study III, the investigation becomes correlational rather than experimental, and the question of an artifact arises again with renewed cogency. Second, is the relationship between test-anxiety and blame-assignment in test and school situations a specific instance of a more general relationship between test-anxiety and attitudes of aggression toward self and others? This question arises from the results of the school dissatisfaction questionnaire in Study II, and it is also suggested by the findings of Sarason and Mandler (1952) in their correlations of test-anxiety and the aggressive subscales of the Waterhouse and Child (1953) questionnaire on "habitual reactions to frustration." If this question is to be answered in the affirmative, then the relationship between test-anxiety scores and blame-assignment takes on a new perspective.

To shed some light on the first of these problems, it was decided to determine if a blame-assignment relationship to test-anxiety scores could be found in a medium other than that of questionnaires. To effect this, use was made of a modification of the thematic apperception technique. If a relationship between test-anxiety scores and amount of blame-assignment for failure in the stories could be established, it would lend support to the belief that we are dealing with a relationship between test-anxiety scores and attitudes toward failure in school situations that exist independently of any artifact introduced by a questionnaire technique of measurement.

The scope of the present investigation permitted only a very circumscribed attack on the second problem. While recognizing the limitations of such an approach, it was decided to use the Children's Form of the Rosenzweig Picture-Frustration study as a measure of aggressive attitudes in other than test situations.

METHOD

Four fifth-grade and three sixth-grade classes were administered the blame-assignment questionnaire in regular class sessions. From each class *Ss* were selected

for individual testing on the basis of the availability of previous test materials and background information which would be relevant to the present study. Thus, only those *Ss* were considered for testing with the projective instruments who had on file with the Children's Test-Anxiety Project the following information: test-anxiety scores from 1954 and 1955, arithmetic aptitude scores from 1955, and a negative history of referral to the school psychologist for behavioral problems. While such information was available for the majority of *Ss*, it varied from class to class. In some classes complete, relevant information might be available for only 30% of the class; in other classes, it reached 90% or more. The presence or absence of this information was assumed due to chance factors, e.g., a redistribution of the school population in Milford in 1955 because of the opening of several new schools which produced a reshuffling of the original test-anxiety sample of 1954.

From the *Ss* for whom complete test and background information was available, it was decided to exclude those whose test-anxiety scores showed too marked a variability from 1954 to 1955. The decision as to what constituted too marked a variability was purely arbitrary. Using the median scores of the total sample given the test-anxiety questionnaire in 1954 and 1955, it was decided to exclude *Ss* who had made marked jumps in scores from one side of the median in 1954 to the other in 1955. Thus, any change in test-anxiety scores which was greater than 10 points and which included a change from one side of the median to the other was disallowed. Since only a limited number of *Ss* could be tested in individual sessions, and since the test-anxiety scores to be used in the correlation were a year old, it was felt that discarding those who had showed marked variability in test-anxiety scores would increase the reliability of any correlation obtained between test-anxiety and blame-assignment.

In all, 62 *Ss* were selected for individual testing, 20 boys and 20 girls from the fifth grade and 11 boys and 11 girls from the sixth. The blame-assignment questionnaire in the same form as in Study III was administered to each class as a group. The instructions told the *Ss* that *E* was interested in finding out how children felt about arithmetic tests and what kept them from doing their very best in arithmetic. The *E* informed the children that in seeking this information he hoped eventually to help other children in their difficulties with arithmetic. The children were also assured that their answers to the questionnaires would be seen only by *E*. After filling out the blame questionnaire, *Ss* were given another questionnaire which related to their experience of anxiety in relation to arithmetic. This instrument was included because the blame-assignment scale was concerned with arithmetic rather than test situations in general. This scale had the same format as the blame-assignment scale and consisted of 16 pairs of items such as the following:

- 5a. When I am called on in class for arithmetic, I sometimes feel my heart beat faster.
- 5b. When I am called on in class for arithmetic, I never feel my heart beat faster.
- 9a. When I am doing arithmetic homework I don't worry about my answers.

- 9b. Sometimes when I am doing arithmetic homework, I worry that I will not have the right answers.

Finally, the children were told that during the following week several of them would see *E* in individual sessions and that the purpose of these sessions would be for *E* to collect information that would better enable him to understand their attitudes toward arithmetic.

In these individual sessions, each *S* was again told that the purpose of the study was to find out how children felt about arithmetic and about other things which might help *E* to understand their attitudes toward arithmetic. They were reassured that what they said or did in the session would be known only to *E* and that it would in no way affect their school work or grades. After this general orientation, *S* was given a Picture-Frustration test blank and the standard instructions, with the exception that they were asked to verbalize their responses rather than to write them. Responses were recorded by *E*. After the Picture-Frustration test, *S* was presented with eight pictures, one at a time. The situation in each picture was described by *E* as one in which a child in the picture had just undergone some failure in arithmetic, either in a class recitation, a test, or on a report card. The *S* was requested to make up a story about the child in the picture which would tell how his failure occurred, what he was thinking and feeling, and what happened afterwards. Six of the pictures were taken from the Michigan Picture Test [3, 4B, 8G, 10B, 10G, 11G], and two were reproductions of photographs that had appeared in the *Journal of the National Education Association*. In all pictures, a child was shown in either a school situation or alone in the presence of an adult who was represented to *S* as either a principal or a parent dissatisfied with the child's performance in arithmetic.

RESULTS

Analysis of scales. Each *S* had self-blame and other-blame and arithmetic-anxiety scores obtained from the questionnaires administered to each class. In addition, arithmetic-aptitude scores were available from school records. In the fall of 1955, all fifth-grade youngsters were given the Iowa Every-Pupil Tests of Basic Arithmetic Skills (Advanced Battery, Test-D, Form P), and all sixth-grade youngsters were given the Arithmetic Fundamentals and Arithmetic Problem sections of the Metropolitan Achievement Test (Intermediary Battery, Form R). Scores from these tests were normalized to permit a combining of both fifth and sixth grades in the analysis of the data. All other scores in the investigation were treated in their raw forms. Since the 1955 test-anxiety score was the most recent test-anxiety score, it was decided to use it in preference to the 1954 score, and all test-anxiety correlations reported here are based on the 1955 scores.

Intercorrelations were obtained for boys and girls and for the fifth and sixth grades. Since no consistent patterns of differences were noted, and since distributions of scores for the subgroups were similar, the intercorrelations are presented in Table 1 for the group of 62 Ss as a whole.

As expected on the basis of the earlier studies, test-anxiety correlates positively with self-blame (.577, $p < .01$) and other-blame (.359, $p < .01$). Arithmetic-anxiety also correlates positively with self-blame (.264, $p < .05$) and other-blame (.510, $p < .01$). The two anxiety scales, test-anxiety and arithmetic-anxiety, correlate positively (.365, $p < .01$). Likewise, the two blame scales correlate positively (.403, $p < .01$). Arithmetic aptitude correlates negatively with test-anxiety ($-.322$, $p < .05$), a result which is in agreement with previous findings of Sarason et al. (1952; in press) on the relationships between test-anxiety scales and achievement scales. There are no significant relationships between arithmetic aptitudes and arithmetic-anxiety or the blame scales.

Since arithmetic aptitude might be conceived to be a primary factor determining S's attitudes of blame-assignment in relation to arithmetic, it was decided to partial arithmetic aptitude out of the test-anxiety and blame correlations. The results are presented in Table 2. The partial correlations of test-anxiety and self-blame with arithmetic constant (.546, $p < .01$) and test-anxiety and other-blame with arithmetic constant (.358, $p < .01$) remain significant. Likewise, the partial correlation of arithmetic-anxiety and other-blame with arithmetic constant (.514, $p < .01$) remains significant. But the partial correlation

TABLE 2
PARTIAL CORRELATIONS OF ANXIETY AND
BLAME SCALES WITH ARITHMETIC
CONSTANT
($N = 62$)

	Self-blame	Other-blame
Test-anxiety	.546*	.358*
Arithmetic-anxiety	.236	.514*

* Significant at .01 level.

of arithmetic-anxiety and self-blame with arithmetic constant (.236) fails to reach the .05 level.

Analyses of the stories. Analyses of the stories for blame-assignment were made in two different ways. In the first method, three independent judges rated the stories for hero-blame assignment and other-blame assignment. Each judge was presented with all 62 stories to a given picture and asked to place them in four piles of increasing amounts of blame assigned to the hero of the story. The judges were asked not to infer hero-blame on the basis of their knowledge of psychological dynamics but to be concerned only with the overt expression of hero-blame in the story. After the judge had placed the 62 stories for the first picture into piles numbered 1 through 4 in order of increasing hero-blame, he was given the 62 stories for the next picture, and so on, until all stories for all Ss had been categorized. At no time was the judge aware of the names or test-anxiety scores of the storytellers. The sum of the categories assigned to all eight stories of a given S by the three judges was considered the hero-blame assignment score for that S. The reliability of the hero-blame judgments was determined by Hoyt's (1941) analysis of variance technique for test reliability. Stories were arranged in columns and subjects in rows with the combined judges' ratings forming the cell entries. The reliability coefficient so obtained was .69. After the judges had made their hero-blame categorizations of the stories, they were asked to go over them again to judge the amount of other-blame assignment in the stories, i.e., blame of teacher, parents, other children, etc., for the hero's failure. The judgments were made in the same manner as for hero-blame. Stories were placed in four piles of increasing other-blame, and the sets of stories to the various pictures were judged separately.

TABLE 1
CORRELATIONS OF ANXIETY AND BLAME
SCALES AND ARITHMETIC APTITUDE
($N = 62$)

	Arithmetic-Anxiety	Self-blame	Other-blame	Arithmetic Aptitude
Test-anxiety				
Arithmetic-anxiety	.365**	.577**	.359**	-.322*
Self-blame		.264*	.510**	-.226
Other-blame			.403**	-.230
				-.137

* Significant at the .05 level.

** Significant at the .01 level.

TABLE 3

CORRELATIONS OF ANXIETY SCALES WITH JUDGMENTS OF HERO-BLAME AND OTHER-BLAME AND NUMBER OF HERO-BLAME, OTHER-BLAME AND NONBLAME STATEMENTS

($N = 62$)

	Hero-blame Judgments	Other-blame Judgments	Hero-blame Statements	Other-blame Statements	Non-blame Statements
Test-anxiety	.252*	.053	.345**	-.018	-.026
Arithmetic-anxiety	-.202	.155	-.103	.006	-.126

* Significant at .05 level.

** Significant at .01 level.

Again, overt rather than inferred blame was to be judged. The categorizations of all three judges for the stories of a given S were combined to give an other-blame assignment score for that S .

The correlations of these judgments of hero-blame and other-blame with test-anxiety and arithmetic-anxiety are presented in Table 3. The correlation of test-anxiety and hero-blame (.252, $p < .05$) is significant,⁶ but that of test-anxiety and other-blame (.053) is not. While arithmetic-anxiety shows no significant correlations with either hero-blame or other-blame, the negative correlation of arithmetic-anxiety and hero-blame (-.202) is of especial note in that it is in a direction opposite to that which would be expected on the basis of the positive correlation between test-anxiety and arithmetic-anxiety (.365).

The second method of analysis of the stories consisted of breaking the stories down into grammatical units. Every sentence or independent clause was considered a unit, and these units were categorized by the judge who formed them as being either statements of hero-blame, other-blame, or nonblame. The judge who made these categorizations had participated in the judging of the stories for hero and other-blame as described above. An interval of several weeks had elapsed between her two tasks, however, and she was still unaware of the anxiety ratings of the S s. From these categorizations of grammatical units, it

⁶ Correlations for test-anxiety and hero-blame judgments were also calculated separately for each judge, using the sum of each judge's hero-blame categorizations for an S as the hero-blame score. The correlations for these judges were respectively .274 ($p < .05$), .265 ($p < .05$), .158.

was possible to obtain hero-blame, other-blame, and nonblame scores for each S simply by adding up the number of statements in his stories that were classified in each of the three categories. These scores were then correlated with test-anxiety and arithmetic-anxiety and are presented in Table 3. Text-anxiety and hero-blame (.345, $p < .01$) was the only significant relationship. Since the total of hero-blame, other-blame, and nonblame statements for an S can in this analysis be considered a measure of story length, it is apparent that the correlation of hero-blame statements and test-anxiety is not an artifact of over-all story length. If such were the case, nonblame statements, other-blame statements, or both would show positive correlations with test anxiety; in fact, they show negligible, negative correlations.

As in the case of the blame questionnaires, it was decided to partial out arithmetic aptitude from the correlations of the anxiety scales and hero-blame assignment. The results are presented in Table 4. The partial correlations of test-anxiety with judgments of hero-blame (.270, $p < .05$) and with statements of hero-blame (.347, $p < .01$) remain significant. The partial correlations of arithmetic-anxiety and judgments of hero-blame (-.204) and statements of hero-blame (-.118) do not reach significance. Since the underlying assumption in the use of the story technique was that hero-blame would reflect the same attitudes as the self-blame questionnaire, it was thought advisable to correlate the hero-blame scores with the self-blame scale. However, neither the self-blame and hero-blame judgments correlation (.071) nor the self-blame and hero-blame statements correlation (.117) was significant.

Analysis of the P-F test. The Rosenzweig Picture-Frustration protocols were scored according to the directions in Rosenzweig's manual. Only Rosenzweig's *direction of aggression* scores were considered. Although Rosenzweig occasionally permits the scoring of a single response in more than one category, this practice was not followed in the present study.

Each S was assigned extrapunitive (E), intropunitive (I), and impunitive (M) scores, which represented the number of responses that he made in each of the respective categories. Since the Picture-Frustration test

TABLE 4

PARTIAL CORRELATIONS OF ANXIETY SCALES WITH
HERO-BLAME JUDGMENTS AND STATEMENTS WITH
ARITHMETIC APTITUDE CONSTANT
($N = 62$)

	Hero-blame Judgments	Hero-blame Statements
Test-anxiety	.270*	.347**
Arithmetic-anxiety	-.204	-.118

* Significant at .05 level.

** Significant at .01 level.

contains 24 cartoon situations to which S gives a response, and each response is categorized as either an E , I , or M , these scores are not independent. To determine if the anxiety scores correlated with any or all of them, it was only necessary to correlate test-anxiety and arithmetic anxiety with the sum of the E and I scores and the difference of the E and I scores for each S . $E + I$ and $E - I$ are independent of each other, and if either E , I , or M is correlated with test-anxiety or arithmetic-anxiety, there should be a correlation between the anxiety score and $E + I$ and/or $E - I$. This analysis gave no evidence of a statistically significant correlation between test-anxiety and $E + I$ ($-.011$) or $E - I$ ($.015$) or between arithmetic-anxiety and $E + I$ ($.163$) or $E - I$ ($.090$).

DISCUSSION

The general hypothesis of this investigation was that in a failure situation, high-test-anxious S s assign blame for a failure to themselves to a greater extent than do low-test-anxious S s. Specifically, we wished to see if this relationship, which had been found with college S s with forced choice between self-blame and other-blame responses, could be obtained with an elementary school population when choice was free for self-blame, other-blame, or nonblame responses. The over-all results of the investigation tend to support the hypothesis, but with qualifications that immediately change the perspective with which one must view the problem of test-anxiety and blame-assignment.

First, the correlations of test anxiety and hero-blame give some support for the contention that we are not dealing with a relationship that is solely a product of the questionnaire

technique of measurement. If S tends to identify with the story hero represented in a failure situation, then his attitudes of blame-assignment expressed in the stories should reflect S 's own tendencies toward blame-assignment in a failure situation. However, the contention that the correlation of test-anxiety and self-blame receives support from the test-anxiety and hero-blame correlations must be tempered by a consideration of the self-blame and hero-blame correlations. Neither the self-blame and hero-blame judgments ($.071$) nor the self-blame and hero-blame statements ($.117$) are significantly correlated. A possible explanation of this state of affairs is that both self-blame and hero-blame tap an underlying aggressive attitude toward the self but that the differences in their modes of expression account for their failure to correlate. Thus, a highly test-anxious S who will not or cannot directly express a hostile attitude toward himself in the self-blame questionnaire may express such an attitude when expression is possible through the more indirect device of a projective technique or vice versa. Again, the projective technique may complicate the expression of the underlying aggressive attitude by its dependence upon the fluctuation in the degree of identification with the hero from S to S and from story to story. It is also possible that the expression of such concomitant attitudes as guilt or shame in the stories may complicate and confuse the direct expression of the self-aggressive attitude. These factors would all tend to lower a self-blame and hero-blame correlation without necessarily affecting the relationship of the underlying self-aggressive attitude with test-anxiety.

The failure to obtain a correlation between test-anxiety and other-blame-assignment in the stories poses another problem. If the relationship between test-anxiety and self-blame obtained in the questionnaires is supported by the test-anxiety and hero-blame correlation, why is not the test-anxiety and other-blame correlation, which is also found in the questionnaires, supported by the correlation of test-anxiety and other-blame-assignments in the stories? It may be that the story technique is not well suited to eliciting responses of other-blame from S . An examination of the number of grammatical units assigned to the hero-

blame category as against those assigned to other-blame revealed means of 11.0 hero-blame units and 1.7 other-blame units for all *Ss*. There is also the possibility that the relationship between test-anxiety and self-blame or hero-blame is just more prepotent than that of test-anxiety and other-blame. Support for this type of inference may be found in the fact that for college *Ss*, faced with a choice between self-blame and other-blame responses, it was the self-blame responses which correlated positively with test-anxiety scores. Then, there is also the question of *S*'s interpretation of the permissiveness of the other-blame response. In the questionnaire *E*, by presenting *S* with the other-blame response already written out, may by implication be adopting a permissive attitude toward the other-blame response which permits *S* to use it. But in the story there is less obvious permission to use the other-blame response; hence, its use may be inhibited. In this question of an artifact of the questionnaires as responsible for the correlation of test-anxiety and self-blame, we must also bear in mind the results of the college study, in which the test-anxiety and self-blame relationship was produced under one—apparently the more stressful—condition of administration but not under the other. Such results suggest that the relationship can be manipulated in a failure situation, and if the relationship can be experimentally controlled, the question of a correlational artifact disappears. Unfortunately, the present investigation leaves the control of the test-anxiety and blame relationship in the case of children still open to question.

Another question raised the possibility that the relationship between test-anxiety and blame-assignment is a specific instance of a higher order relationship between test-anxiety and aggressive attitudes in general. This notion receives no support from results with the Rosenzweig Picture-Frustration test. While this is not to say that the possibility of a higher order relationship is ruled out, it does suggest that the relationship between test-anxiety and aggressive attitudes is not a completely general one and that there must be some tendency toward specificity. Whether the specificity limits itself to test situations and related school experiences we are not in a position to

say. But to the degree that it does show a tendency to specificity, we are justified in finding in it a problem for the ultimate understanding of the nature of test-anxiety and its relationship to intellectual performance and school adjustment.

Finally, there is the possibility that specific failure had no effect on blame-assignment, because the cumulative effect of past experience with arithmetic could conceivably so deeply ingrain attitudes in *Ss* concerning arithmetic that one additional failure cannot affect them. To this possibility, this investigation offers no direct support or rebuttal. One might conceivably argue that every *S* has already experienced considerable failure in arithmetic. Thus, in a sense, he is already in a failure situation and when presented with the opportunity to assign blame for his performance, he proceeds to do so in accordance with the degree of anxiety he experiences in test situations without regard to any additional failure which *E* may briefly impose. Such an argument cannot be rejected by our study, but, on the other hand, it cannot be supported by it. Nevertheless, this *post hoc* explanation may prove to be of eventual use in understanding the dynamics of the relationship between test-anxiety and blame-assignment. And this possible explanation must be taken into account in planning any future attempts to experimentally control the test-anxiety and blame-assignment relationship.

While the main focus of this study has been on the relationship of blame-assignment to the test-anxiety scale as developed by Sarason, the inclusion of a specially devised arithmetic-anxiety scale permits some interesting comparisons. The two anxiety scales differ in format, but have a fair amount of overlap in content, and they purport to measure closely allied phenomena. In fact, the only essential difference is that arithmetic-anxiety measures a specific type of test-anxiety, whereas the test-anxiety scale taps a larger range of test situations in which anxiety may be invoked. Their similarity leads us to expect similar relationships with the other measures employed in this study, and, to a large extent, this expectation was borne out. Like test-anxiety, arithmetic-anxiety correlated positively with the self-blame and other-blame scales, but did not

correlate with the Picture-Frustration measures. However, the test-anxiety scale has a significant positive correlation with judgments of hero-blame, whereas the arithmetic-anxiety scale has an insignificant negative correlation with hero-blame.⁷ Again, closer examination of the questionnaire results show that whereas test-anxiety and self-blame (.577) correlated more highly than test-anxiety and other-blame (.359), arithmetic-anxiety and self-blame (.264) correlated lower than arithmetic-anxiety and other-blame (.510).⁸ It would thus seem that while test-anxiety and arithmetic-anxiety share considerable communality, there must still exist important areas of difference between the two scales. The sources of this difference may lie in any one of several places. The formats of the anxiety scales are different, and in relation to the blame questionnaires and stories they were administered at different times and in different orders. It would be hazardous to guess at the effect of these differences, but before one could be certain of a genuine difference between the two anxiety scales, these factors would have to be ruled out. However, there is the possibility that the difference between the scales derives from their difference in specificity. The arithmetic-anxiety scale concerns anxiety in the very area in which *S* is asked to assign blame, i.e., arithmetic. While test-anxiety undoubtedly includes the arithmetic situation, it

⁷ The significance of this difference between test-anxiety (TA) and arithmetic anxiety (AA) in correlation with hero-blame (HB) can be demonstrated by use of Quenouille's (1952) formula for the correlation of a variable with the difference between two other variables:

$$r_{HB(AA-TA)} = \frac{r_{HB(AA)} - r_{HB(TA)}}{\sqrt{2(1 - r_{AA(TA)})}} = -.403(p < .01)$$

⁸ The significance of the difference between these two sets of correlations can be inferred from the significance of the correlation of the difference between arithmetic-anxiety (AA) and test-anxiety (TA) with the difference between other-blame (OB) and self-blame (SB). This may be obtained by a logical extension of Quenouille's (1952) formula:

$$\begin{aligned} r_{(AA-TA)(OB-SB)} &= \frac{(r_{AA(OB)} - r_{AA(SB)}) - (r_{TA(OB)} - r_{TA(SB)})}{\sqrt{4(1 - r_{OB(SB)})(1 - r_{AA(TA)})}} \\ &= .377(p < .01) \end{aligned}$$

includes much more. The specificity of the anxiety to the area of blame assignment may therefore account for the difference. One may speculate rather freely that the more specific the anxiety is to the area in which blame is assigned, the more prepotent is the other-blame response, and the less prepotent is the self-blame response. If such an explanation were allowed, it would point out the desirability for specifying as closely as possible the stimulus for anxiety responses in the design of anxiety scales. We cannot expect the performance or the attitudes of *S* to remain constant in different situations if *S*'s level of anxiety varies with the situations. Thus, while a test-anxiety scale may be a more suitable instrument than a general anxiety scale for a study of the effects of anxiety in a test or performance situation, certain problems may require us to specify the stimulus that sets off *S*'s anxiety response with even greater precision.

SUMMARY

In this investigation an attempt was made to measure the relationship between test-anxiety in elementary school children and the propensity for self-blame assignment in regard to the experience of failure in a test situation. Four related studies were conducted with fifth- and sixth-grade school children. Test-anxiety scores were available for the *Ss* from a previous administration of the CTA scale. In these four studies, *Ss* filled out questionnaires that permitted them to assign blame to themselves or others for their performance in arithmetic. In Studies I and II these questionnaires were administered after an induced failure in arithmetic. In Study III, the blame questionnaire was administered both before and after the failure in arithmetic, and in Study IV the blame questionnaire was administered without the induction of a specific failure in arithmetic. In all of these studies the questionnaires were administered in group situations. However, in Study IV the group situation was followed by individual testing of the *Ss* with projective instruments.

The results are interpreted as indicating that although test-anxiety and self-blame are related in a failure situation, the relationship, at least in this case of an arithmetic test, may be shown to exist independently of a specif-

ically induced failure. While the possibility of a correlational artifact is not definitely eliminated, the existence of a test-anxiety and blame relationship in the thematic apperception technique and the previous study with college Ss suggests that the relationship is not artifactual. Although test-anxiety correlated with the other-blame questionnaire, no support was found for this relationship in the thematic apperception technique. The use of an arithmetic-anxiety scale seemed to support generally the findings with the test-anxiety scale. However, some differences between arithmetic-anxiety and test-anxiety in their relationships to the various blame scales were pointed out, indicating that although arithmetic-anxiety and test-anxiety share considerable communality, they yet have significant areas of difference in their relationships to the blame scales.

REFERENCES

- CHILD, I. L. Personality. *Annual Rev. Psychol.*, 1954, 5, 149-170.
 DOLLARD, J., MILLER, N. E., MOWRER, O. H., SEARS,

- R. R., FORD, C. S., HOVLAND, C. I., & SOLLENBERGER, R. T. *Frustration and aggression*. New Haven: Yale Univer. Press, 1939.
 DORIS, J. Test-anxiety and blame-assignment in grade school children. Unpublished doctoral dissertation, Yale Univer., 1957.
 DORIS, J., & SARASON, S. Test anxiety and blame assignment in a failure situation. *J. abnorm. soc. Psychol.*, 1955, 50, 335-338.
 FREUD, A. *The ego and the mechanisms of defense*. New York: International Universities Press, 1946.
 HOYT, C. Test reliability estimated by analysis of variance. *Psychometrika*, 1941, 6, 153-160.
 QUENOUILLE, M. H. *Associated measurements*. London: Butterworths Scientific Publications, 1952.
 SARASON, S. B., DAVIDSON, K., LIGHTHALL, F., & WAITE, R. A test anxiety scale for children. *Child Developm.*, in press.
 SARASON, S. B., & MANDLER, G. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, 47, 810-817.
 TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, 53, 303-320.
 WATERHOUSE, I. K., & CHILD, I. L. Frustration and the quality of performance: III. An experimental study. *J. Pers.*, 1953, 21, 298-311.

Received November 1, 1957.

EGO DISJUNCTION AND PSYCHOPATHOLOGY¹

ARNOLD TREHUB

V.A. Hospital, Northampton, Massachusetts

THIS paper is concerned with the concept of ego disjunction and its relationship to severity of psychopathology. Ego disjunction is defined here as the concurrence in an individual of two or more relatively long enduring needs or characterological dispositions with mutually incompatible objectives and with a relatively high level of joint strength. According to this definition, for example, need Autonomy and need Abasement, as described by Murray (1938) would be considered disjunctive if the joint strength of these needs were high within an individual.

While the implications of antagonistic characterological dispositions have been of concern to clinicians and personality theorists (Cameron & Magaret, 1951; Dollard & Miller, 1950), the concept, to the writer's knowledge, has not been experimentally investigated in its relationship to psychopathology. Since ego disjunction may be considered as a particular kind of intrapsychic conflict, one might assume it to be positively related, over a wide range, with degree of psychopathology. It is this assumption which is investigated here.

It was predicted that if measures of ego disjunction were obtained from college students, adolescents, neurotics, character disorders, and schizophrenics, these measures would describe an ascending monotonic course over the various groups arranged in the order listed above. This order was arrived at by ranking on the basis of the relative degree of maladjustment or psychopathology assumed to be characteristic of each group. The positions of college students, neurotics, and schizophrenics relative to each other is in accordance with widely accepted notions about psychopathology in these groups. Because of the conflict and intrapsychic turmoil commonly attributed to adolescents (Ausubel, 1954; Noyes, 1949) this group was assumed to be

higher on the continuum of psychopathology than college students; while the fact that adolescents as a group are not generally considered to be ill in psychiatric terms suggested that they be ranked below neurotics. Character disorders were ranked above neurotics because of their resistance to treatment, the chronic nature of the distortions in their ego functions, and the frequency with which episodic psychotic behavior is observed in such individuals (Cleckley, 1941; Fenichel, 1945; Noyes, 1949). The experimental hypothesis may be stated: With respect to degree of ego disjunction, College Students < Adolescents < Neurotics < Character Disorders < Schizophrenics.

METHOD

The subjects (Ss) were 100 males, with 20 in each of the following five groups: College Students, mean age = 23.8 years; Adolescents, mean age = 15.6 years; Neurotics, mean age = 34.2 years; Character Disorders, mean age = 35.4 years; Schizophrenics, mean age = 31.0 years. The classification of Neurotics, Character Disorders, and Schizophrenics was based upon established psychiatric diagnoses. With the exception of six Ss with character disorders, all psychiatric Ss were hospitalized at the time of testing.

Measures of ego disjunction were obtained by the following procedure. Four pairs of needs in the Edwards Personal Preference Schedule (Edwards, 1954) were selected by the investigator as disjunction indicators on the basis of apparent mutual incompatibility of objectives within each pair. The need pairs selected are listed below.

Aggression-Deference
Succorance-Nurturance
Autonomy-Abasement
Order-Change

Other pairs displayed apparent incompatibility but were excluded from the list either because the mutual incompatibility did not appear to be of sufficient degree, or because inclusion would have resulted in several needs appearing more than once in the scale. For purposes of this study, it was decided to avoid duplicate scoring of needs despite the fact that a need might enter into several disjunctive combinations.

Having chosen need pairs which may indicate disjunction, one is faced with the problem of deciding how strong these needs must be before one wishes to assert that ego disjunction is present. The approach used in this study was to evaluate the joint magnitude of the needs within the indicator pairs. It was arbitrarily decided to sum the two need scores (expressed in standard score units) for each disjunction indicator, and to

¹ The author is indebted to James Klett and Mack Knutsen for having made available some of the EPPS records used in this study, and to Isidor Scherer, Cesareo Pena, Arthur Tamkin, James Klett, and Henry Oppenheim, and Louis Nidorf for their constructive review of the manuscript.

subtract 100 from each sum.² Positive residuals were considered as indicating the presence of disjunction, and the degree of disjunction for each indicator pair was considered to be reflected in the magnitude of the positive residual. The total ego disjunction score for each *S* consisted of the sum of his positive residuals on the disjunction indicators.

RESULTS

Mean ego disjunction scores for each of the five groups are shown in Fig. 1. With the exception of the finding that Adolescents scored at the level of Neurotics, the obtained pattern and the results of a test of variance between groups support the experimental hypothesis ($F = 8.30, df = 4$ and $95, p < .001$).

After the over-all null hypothesis was tested, the data were further analyzed by the application of Tukey's test for significant gap (Edwards, 1954). The significant gap was set at $p = .05$ and a one-tail t value was used. By this procedure, three subgroups based upon differences in ego disjunction were distinguished within the original five groups. These subgroups are (a) Schizophrenics; (b) Character Disorders, Neurotics, Adolescents; and (c) College Students. When differences between individual means were tested, it was found, as shown in Table 1, that there were no significant differences among means within the subgroup of Character Disorders, Neurotics, and Adolescents.

The foregoing analyses were concerned with the magnitude of total ego disjunction scores obtained by summing over four disjunctive need pairs for each *S*. In order to determine if there were significant differences among the groups with respect to particular disjunctive pairs, Kolmogorov-Smirnov tests (two-tail) (Siegal, 1956) were applied to the differences between groups for each disjunctive pair. It was necessary to use a nonparametric analysis in this instance because scores were markedly skewed when disjunctive pairs were considered separately. Of 40 tests carried out, only the difference between Schizophrenics and College Students on the need Succorance-need Nurturance pair reached significance ($p = .05$). Since one difference at the .05 level can be expected

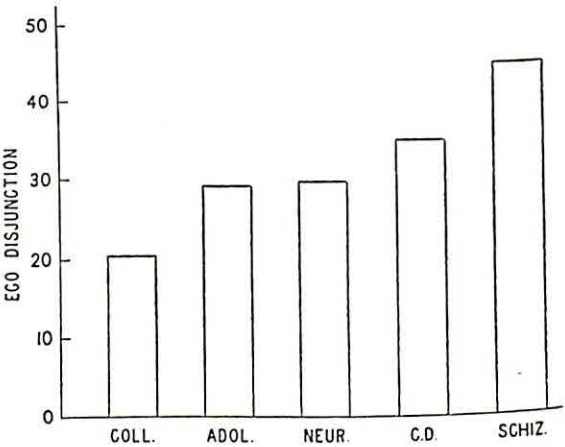


FIG. 1. MEAN EGO DISJUNCTION SCORES FOR COLLEGE STUDENTS, ADOLESCENTS, NEUROTICS, CHARACTER DISORDERS, AND SCHIZOPHRENICS

TABLE 1
DIFFERENCES BETWEEN MEAN MEASURES OF EGO DISJUNCTION (ABOVE DIAGONAL), AND t VALUES (BELOW DIAGONAL)

Group	College Students	Adolescents	Neurotics	Character Disorders	Schizophrenics
College Students	—	8.75	9.00	14.05	23.85
Adolescents	2.01*	—	.25	5.30	15.10
Neurotics	2.07*	.06	—	5.05	14.85
Character Disorders	3.23**	1.22	1.16	—	9.80
Schizophrenics	5.48**	3.47**	3.41**	2.25**	—

* $p < .05$.
** $p < .005$.

to occur by chance alone when 40 tests are made, we are unable to characterize any of the groups with respect to particular areas of disjunction.

DISCUSSION

It will be observed that approximate equality of strength of antagonistic needs was not set as a condition within the definition of ego disjunction. The reason for not demanding that antagonistic needs be at the same level follows from an analysis of the dynamics of the concept.

Any disjunctive pair of needs, as defined in this study, can be represented as a double approach-avoidance conflict (Dollard, 1950) in which the level of approach for each need (i.e., the score obtained for that need) is also a measure of the level of avoidance for its

² In the EPPS normative sample, the score of 50 was set as the mean standard score for each need. Thus the value of 100, which is subtracted from the sums of the disjunction indicator pairs, represents the mean standard score multiplied by two.

paired opposite need. In considering the issue of the relative parity of antagonistic need levels as it relates to degree of intrapsychic conflict, the following question is central: Where along the course of the approach-avoidance gradients do we obtain our measures of antagonistic needs when we assess need strength in a test situation using an instrument such as the EPPS? Certainly, the answer to this question cannot be given in terms of a precise point, but the writer believes it safe to assume that in the paper and pencil test situation, there is generally a substantial psychological distance between the *S* and the actual goal responses represented by the needs measured at the moment. This distance between *S* and goal, which is assumed to exist in the test situation, establishes the possibility of further movement in the direction of the goal in other situations.

Now let us imagine that we have obtained the test scores of a pair of antagonistic needs for two different individuals, that the *joint* strength of the needs is the same for each of the individuals, but that for Individual A the needs within the pair are equal in strength, while for B they are unequal in strength. In the case of A we would expect, other things being constant, that he would neither approach nor withdraw from the goal responses represented by his measured needs. Thus, according to theory, his total conflict would remain relatively stable at the level indicated by the joint strength of his antagonistic need scores. Individual B, on the other hand, would tend to approach the goal responses represented by his stronger need and, because of the relationship between approach and avoidance gradients, conflict would occur at a relatively more intense level. By the same theoretical principles, however, he would tend to withdraw from the goal responses represented by the weaker need, with the result that he would experience this conflict at a relatively less intense level. The increment in the intensity of one conflict and the decrement in the intensity of the other sum to zero (assuming linear gradients). The net result is that total conflict for B, as for A, is at the level indicated by the *joint* strength of his antagonistic need scores despite the disparity in his scores.

The foregoing theoretical considerations suggest that parity of antagonistic need

strengths should not be demanded within a definition of ego disjunction.

From an empirical point of view, if relative parity of antagonistic needs were positively related to intensity of intrapsychic conflict, we would expect differences in strength between such needs to be less among poorly adjusted individuals than among relatively well-adjusted individuals. When the data collected in this study were analyzed, however, no evidence was found to indicate such a relationship. On the contrary, when College Students were compared with Schizophrenics, it was found that the Schizophrenics showed somewhat greater disparity of antagonistic need scores (mean disparity for College Students = 10.21; mean disparity for Schizophrenics = 13.51).

It should be emphasized that the disjunctive need pairs measured in this study do not constitute the only possible areas of ego disjunction. Within the framework of the EPPS alone, other disjunctive pairs can be composed if certain needs are entered into more than one combination. Furthermore, needs not measured by the EPPS may form disjunctive combinations. The scale constructed for the purpose of this investigation represents a sample of what the writer believes are some of the important areas of ego disjunction.

The increase of ego disjunction score values over the five groups tested supports not only the hypothesis investigated, but also the applicability of the scale used to measure ego disjunction. Preliminary observations suggest that it might prove to be valuable as a clinical instrument as well as for research applications.

SUMMARY

This study was designed to test the hypothesis that degree of ego disjunction is positively related to degree of psychopathology. A scale based upon the Edwards Personal Preference Schedule was devised to measure ego disjunction. College Students, Adolescents, Neurotics, Character Disorders, and Schizophrenics were tested and the following results were found.

1. The pattern of the obtained ego disjunction scores corresponded in general to the predicted pattern. Thus a positive relationship between degree of ego disjunction and degree of psychopathology was demonstrated.
2. Statistical analysis of ego disjunction

scores yielded three significantly different subgroups. The first subgroup consisted of Schizophrenics alone; the second consisted of Character Disorders, Neurotics, and Adolescents; the third consisted of College Students alone.

3. There were no significant differences among means within the subgroup of Character Disorders, Neurotics, and Adolescents.

4. None of the groups could be characterized with respect to particular areas of ego disjunction.

It was suggested that the ego disjunction scale used in this study shows promise for both clinical and research applications.

REFERENCES

AUSUBEL, D. P. *Theory and problems of adolescent development*. New York: Grune & Stratton, 1954.

CAMERON, N., & MAGARET, ANN. *Behavior pathology*. Boston: Houghton Mifflin, 1951.

CLECKLEY, H. *The mask of sanity*. St. Louis: Mosby, 1941.

DOLLARD, J., & MILLER, N. E. *Personality and psychotherapy*. New York: McGraw-Hill, 1950.

EDWARDS, A. L. *Manual for Edwards' Personal Preference Schedule*. New York: Psychological Corp., 1954.

EDWARDS, A. L. *Statistical methods for the behavioral sciences*. New York: Rinehart, 1954.

FENICHEL, O. *The psychoanalytic theory of neurosis*. New York: Norton, 1945.

MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.

NOYES, A. P. *Modern clinical psychiatry*. (3rd ed.) Philadelphia: Saunders, 1949.

SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.

Received November 18, 1957.

COGNITION, MOTIVATION, AND TIME PERCEPTION¹

PETER SCHÖNBACH

Frankfurt/Main, Germany

IN AN unpublished exploratory study on the effects of social isolation by Schachter, a few student volunteers lived in isolation for several days. These subjects reported that whenever they thought about anything connected with the outside world they would suffer most from being isolated and time seemed to drag, but when they concentrated on objects within the room or "didn't think about anything," the situation was hardly annoying and the time passed quickly.

Conceptualizing the outside world as a goal and being isolated as a barrier situation for these students, one may tentatively conclude from these findings: (a) The more a person thinks about a goal while being in a barrier situation, the greater will be the force acting on him to reach that goal. (b) The greater this force, the greater will be the person's estimate of the time spent in the barrier situation. A test of these two hypotheses requires that they be stated more rigorously. If a person has no need for a goal, no force will act on him to reach it, whether he thinks about it or not. Both need and ideation relevant to the goal seem to be necessary conditions for a goal-directed force to act on a person, and accordingly we may formulate:

1. The force acting on a person during a time interval T to reach a goal G is an increasing function of the person's need for G "times" the relevance of his ideation during T .

2. The greater the magnitude of the force to reach a goal acting on a person in a barrier situation (within limits), the greater will be his estimate of the time spent in the barrier situation.

¹ This paper is based on a thesis offered in partial fulfillment of the requirements for the Ph.D. degree at the University of Minnesota. The research was carried out at the University of Minnesota under sponsorship of the Office of Naval Research (Contract N8onr-66216). The author wishes to thank his adviser Stanley Schachter for his inestimable aid and encouragement at every stage of the study. He also wishes to thank the Laboratory for Research in Social Relations at the University of Minnesota for the support given to this project, and to D. H. Lawrence for some valuable suggestions.

The term "need" is used here to designate the motivational state immediately prior to the experimental manipulation of ideational content, and the term "force" designates the motivational state resultant from the interaction of "need" and the manipulation of ideational content. Relevance of ideation with respect to a goal G is defined as follows: An ideational element that contains the ideational equivalent of G is relevant to G ; an ideational element that does not contain the ideational equivalent of G is irrelevant to G . The greater the proportion of relevant elements in an ongoing ideational process, the higher is its relevance with respect to G .

In an experiment designed to test the first hypothesis, one has to manipulate relevance of ideation by proper stimulation. On the animal level the drive-increasing effect of external stimuli has been explored in the case of hunger and thirst with positive (Calvin & Bicknell, 1953; Danziger, 1951; Greenberg, 1954) and negative (Siegel & MacDonnel, 1954) results. On the human level many investigators (e.g., Brožek, Guetzkow, & Baldwin, 1951; Klein, 1954; Sanford, 1936; Sanford, 1937; Wispé, 1954) have studied the extent to which drive—especially hunger or thirst drive—influences cognitive processes elicited by various external stimuli. However, no experiment with human subjects appears to have been conducted in which cognitive processes were treated as antecedent, and hunger or thirst as consequent variable. At present, Clark's (1954) experiment on the sex motive seems to be the study closest to our problem.

In the extensive literature on time perception, only a few studies on the effects of motivation on time estimation are reported (Falk & Bindra, 1954; Filer & Meals, 1949; Henrikson, 1948; Rosenzweig & Koth, 1933), and their bearing on the second hypothesis is ambiguous because they were performed under very diverse conditions. With the possible exception of Rosenzweig and Koth's (1933) experiment, they seem to be in agreement with it. A discussion of the relevant literature can be found in Schönbach (1956).

EXPERIMENT I

Method

For an initial test of our hypotheses it seemed advisable to look for a "simple" force that could be manipulated and measured with comparative ease, and consequently a hunger setting was chosen. The specific experimental variables corresponding to the three concepts of the general hypotheses were: the need for food, thinking about food, and the force to eat.

Subjects and recruitment. One hundred girls, volunteers from an introductory psychology class at the University of Minnesota, served as Ss in the experiment. Each prospective S was invited by telephone to an "experiment in food tasting." According to a prearranged schedule, E asked 60 of the Ss during the recruitment call not to eat anything on the day of the tasting experiment and to drink only coffee or water (High Need condition). Forty Ss received instructions to eat normally on the day of the tasting (Low Need condition). High Need Ss were scheduled for 3:00 or 4:00 P.M., Low Need Ss for 1:00 or 2:00 P.M.

Design and procedure. Each session was conducted with one S at a time. At the beginning, E once more explained that this would be an experiment on food tasting, which would last at most one and a half hours. In order to standardize expectations he added that S would be served "some luscious chocolate cakes, a variety of cookies, and various kinds of exquisite Viennese pastry." After this, E announced a delay due to some necessary preparations, asked S to lend him her watch under the pretext that his was broken, and created one of the following three stimulation conditions:

1. Relevant Stimulation: He placed before S paper and pencil, a richly illustrated cookbook, and an instruction sheet that directed her to go through some sections of the cookbook ("Desserts," "Pies," "Cakes") and to rate each dish she knew on a five-point scale according to how much she liked it.

2. Irrelevant Stimulation: Besides paper and pencil E handed S a folder with various fashion designs and an instruction sheet that directed her to look at one design after another and to rate each one on a five-point scale according to how much she liked it.

3. No stimulation: No task was assigned to S. At the beginning of the experiment E had seen to it that S, in this condition, did not bring anything that might distract her into the experimental room which was small and empty of any diverting object.

Relevant Stimulation and Irrelevant Stimulation were paired with both High Need and Low Need, No Stimulation only with High Need. Each of these five conditions was run with 20 Ss. They are abbreviated henceforth as follows:

- Hi-Re:* High Need—Relevant Stimulation
- Hi-Ir:* High Need—Irrelevant Stimulation
- Lo-Re:* Low Need—Relevant Stimulation
- Lo-Ir:* Low Need—Irrelevant Stimulation
- Hi-No:* High Need—No Stimulation

The first four conditions are the main ones of the experiment; Hi-No was conducted as a control to learn more about the effects of the stimulations introduced.

After the stimulus manipulation, S was left alone for 13 minutes measured by a stop watch. E then returned with a questionnaire in which S was asked to: (a) estimate the length of the delay period, and to indicate on graphic rating scales (b) how slowly or rapidly time seemed to pass during the delay period, (c) how she liked the rating task (or the waiting), (d) how much she thought about food, (e) how strong her desire was to start with the tasting, (f) whether she felt more or less hungry than at the beginning of the session, and (g) how hungry she was at the moment.

After S had answered these questions, E rolled in a table with an appetizing arrangement of four kinds of cookies labeled A, B, C, D, ten pieces of each kind, and either coffee, tea, or milk, whatever S had chosen during the recruitment call. He told S to eat and drink as much as she wanted and then turned away to read in a book or a journal, because he felt that it might inhibit S if he would watch her eat. After S announced that she had eaten enough, she was asked to indicate on rating scales how she liked the cookies A, B, C, D and how she liked cookies in general.

In the meantime, E removed the cookie table from the room. He then placed a pegboard puzzle before S, together with the rules of this game, and after he had made sure that S understood the task he told her to keep playing until his return and left her alone for another 13-minute period. This puzzle game was added to the core of the experiment in order to obtain time estimates for a period with the same activity for all Ss. After the game period, S received a third questionnaire in which she was asked to estimate the length of this period, to indicate how slowly or rapidly time seemed to pass, and to indicate how she liked working on the game problem. Finally E announced the end of the experiment, explained it in detail and requested S not to talk about it. All Ss appeared to have complied faithfully.

Analysis of the data. Most of the questions had to be answered by marking a continuous scale with six labeled points one inch apart. The distance between a S's mark and an arbitrary zero point was taken as the S's score. For each of the five conditions, means of these scale scores, of the time estimates (converted into seconds), and of the number of cookies eaten were computed. Differences between means were evaluated by two-tail *t* tests. In addition, analysis of variance tests were performed with the data from the four main conditions in order to detect significant interactions of need with relevance of stimulation.

Effectiveness of the manipulations. A comparison of High Need and Low Need conditions (Stimulation held constant) with respect to numbers of cookies eaten and answers to the question "How hungry are you right now?" furnishes an indirect check on the effectiveness of the need manipulation. The mean score on the hunger scale for Hi-Re and Hi-Ir combined is 3.47, the mean for Lo-Re and Lo-Ir combined is 1.08. The difference between these two means is significant at $p < .001$. The mean number of cookies eaten is 7.01 for the two High Need conditions combined and 4.71 for the Low Need conditions combined. This difference is also significant at $p < .001$.

The question "How much altogether did you think

about food. . .^a served as a measure of the effectiveness of the stimulus manipulation. This question was answered by marking a scale which ranged from 0% to 100% of the time. The condition means on this scale were: Hi-Re 91%, Hi-No 43%, Hi-Ir 7%, Lo-Re 88%, Lo-Ir 9%. From a comparison of these means it is clear that not only the differences between Relevant and Irrelevant Stimulation conditions but also the differences between Hi-Re and Hi-No and between Hi-No and Hi-Ir are in the expected direction and very large. All of these differences are significant at a level far beyond $p = .001$. On the other hand, the differences between High and Low Need conditions with stimulation held constant are negligible and not significant.

Considering the results of these tests we may conclude that the conditions necessary for testing the hypotheses were created.

Results and Discussion

Force as a function of Need "times" Relevance of Ideation. The specific prediction derived from the first hypothesis and tested in the experiment was: With need for food held constant at a *high* level, the force to eat acting on a person who is exposed to relevant stimulation will be greater than the force to eat in a person exposed to irrelevant stimulation. With need for food held constant at a *low* level, the force to eat in a person exposed to relevant stimulation will not (or not appreciably) be greater than the force to eat in a person exposed to irrelevant stimulation.

The data on the force to eat consist of the number of cookies eaten and scores from three answer scales on which each *S* indicated after the delay period how strong her desire was to start with the tasting, whether she felt more or less hungry than at the beginning of the experiment, and how hungry she was at the moment. Condition means of these data are reported in Table 1.

On the scale "More or less hungry" there is

a large difference between Hi-Re and Hi-Ir in the expected direction and significant at $p < .001$; the difference between Lo-Re and Lo-Ir is very small and not significant. This result, together with the highly significant ($p < .005$) interaction of need with stimulation demonstrated by an analysis of variance, clearly supports the hypothesis under test. Further evidence for it can be derived from a comparison of Hi-No with both Hi-Re and Hi-Ir. The mean percentage of "thinking about food" in Hi-No was significantly smaller than in Hi-Re and significantly greater than in Hi-Ir. On our measure of the force to eat we should therefore expect the mean score of Hi-No to fall between the means of Hi-Re and Hi-Ir. This is indeed the case, and the two differences, Hi-Re minus Hi-No and Hi-No minus Hi-Ir are accompanied by suggestive levels of significance ($p < .19$ and $p < .07$).

Quite in line with these findings are the results of the question about the level of hunger feelings at the end of the delay period. The difference between Hi-Re and Hi-Ir is large ($p < .01$), the mean for Hi-No falls between those of Hi-Re and Hi-Ir (different from the Hi-Re mean at $p < .03$), and there is practically no difference between Lo-Re and Lo-Ir. The interaction of need with stimulation is significant at $p < .12$.

The data about the "desire to taste" present a slightly different picture. The means of the three High Need conditions stand again in nearly the same relationship to each other as the corresponding means on the other two scales. Hi-Re vs. Hi-Ir is significant at $p < .001$, Hi-Re vs. Hi-No at $p < .01$, and Hi-No vs. Hi-Ir at $p < .11$. In addition, however, the difference between Lo-Re and Lo-Ir is also significant ($p < .02$) in this case, and the mean of Lo-Re is even larger than the Hi-Ir mean, although not significantly so. The finding that stimulation from a cookbook increases the desire to taste in *Ss* with low need, although it does not increase their hunger feelings, suggests that the "desire to taste" is influenced by a factor which we may call curiosity. In line with this interpretation are some comments from Lo-Re *Ss* who had indicated both a strong desire to start with the tasting and feelings of satiation, and who said that they were "just curious" about the tasting. This curiosity

TABLE 1
MEANS OF THE DATA ON THE FORCE TO EAT

Condition	Desire to Taste ^a	More or Less Hungry ^b	How Hungry ^c	Number of Cookies Eaten
Hi-Re	3.23	+0.98	3.68	7.50
Hi-No	2.51	+0.53	3.22	6.93
Hi-Ir	1.93	-0.11	2.98	6.61
Lo-Re	2.26	+0.19	1.13	4.55
Lo-Ir	1.47	+0.10	1.02	4.88

^a The higher the score, the greater the desire to taste.

^b A plus score denotes an increment, a minus score a decrement in feelings of hunger.

^c The higher the score, the greater the level of hunger feelings.

is of course not the only determinant factor of the desire to taste. The large and significant difference ($p < .001$) between Hi-Re and Lo-Re demonstrates that the desire to taste is also a function of the force to eat.

The mean numbers of cookies eaten show almost the same pattern as the means from the "How hungry" question, with only one slight and nonsignificant reversal in the Low Need conditions. The three High Need means are in the same predicted order of magnitude as the corresponding means from the other three measures. In this case, however, the differences are not significant at any conventional level. The test of Hi-Re vs. Hi-Ir resulted in $p < .19$.

It is not very surprising that the differences in consumption among the three High Need conditions are "small" in contrast to the differences obtained with the scales. The Ss answered the questions *before* but, of course, started to eat only *after* the table with the cookies had been placed before them. The enticing arrangement of cookies on this table presented a very real relevant stimulation for all Ss, which may well have had equalizing effects on the forces to eat in the three High Need conditions.

In all five conditions the cookies received very similar ratings after the tasting; differences in attractiveness of the cookies can therefore not be held responsible for the observed differences in the force to eat. Summarizing, we may then conclude that the data just presented lend strong support to the first hypothesis.

The finding that "thinking about food will make a person hungrier" may not seem new to the reader. Yet this fact has not received sufficient attention in the past. It gains special interest from the fact that persons in a state of semi-starvation are excessively preoccupied with thoughts about food (e.g., Brožek & Guetzkow, 1951). Our results, as those of other investigators (cf. Postman, 1953), are at variance with Murphy's hypothesis of autism that ruminations about food are pursued because they have drive-satisfying qualities. They rather suggest that starving prisoners of war, for instance, could minimize their discomfort if they would systematically distract themselves from thinking about food

instead of "daydreaming" about sumptuous meals. A more detailed discussion of these points can be found in Schönbach (1956).

Time estimation as a function of force. The evidence on time estimation for the delay period is based upon three indices: (a) The estimate of the length of the delay period, (b) the discrepancy between delay period estimate and game period estimate, and (c) the judgment about the passage of time during the delay period. Condition means of these three indices are reported in Table 2. In the following paragraphs only the data from the four main conditions will be considered. The average delay period estimate in Hi-No seems to be affected by a gross artifact which will be discussed at the end of this section.

From the two general hypotheses, the following prediction about the *effects of relevance of ideation* can be derived: High Need Ss who were exposed to relevant stimulation should give higher time estimates for the delay period than High Need Ss who were exposed to irrelevant stimulation. Little or no such differences should be expected with Ss in a state of low need.

On both measures, "estimate of the delay period" and "delay period estimate minus game period estimate," the Hi-Re means are much higher than the corresponding Hi-Ir means ($p < .005$ and $p < .04$ respectively). On the other hand, the difference between the mean delay period estimates in Lo-Re and Lo-Ir is comparatively small and significant only at $p < .10$, and on the second index the relationship between these two conditions is reversed, but the difference is small and not significant. Besides, on the difference measure an analysis of variance test demonstrated a significant interaction ($p < .02$) of need with

TABLE 2
MEANS OF THE DATA ON TIME ESTIMATION

Condition	Estimate of the Delay Period (sec.)	Delay Period Estimate minus Game Period Estimate	Passing of the Delay Period ^a
Hi-Re	757	+3	1.56
Hi-Ir	551	-187	1.24
Lo-Re	730	-157	1.27
Lo-Ir	612	-92	1.19
Hi-No	553	-23	2.82

^a The higher the score, the more slowly time seemed to pass.

relevance of stimulation. In short, the data strongly support the derivation just presented.

A second derivation that follows from our two hypotheses is concerned with the *effects of need*: High Need Ss who were exposed to relevant stimulation should give higher time estimates for the delay period than Low Need Ss after relevant stimulation. Little or no such differences should be expected with Ss who were exposed to irrelevant stimulation.

Table 2 demonstrates that in line with this prediction Hi-Re is higher than Lo-Re on both indices under discussion, significantly higher ($p < .06$) only on the second one. The differences between Hi-Ir and Lo-Ir on the two indices are not significant, which is also in agreement with our derivation. Yet on both indices the Hi-Ir mean is lower than the Lo-Ir mean. This finding requires further consideration, and we shall return to it in a later section. The significant interaction of need with relevance of stimulation mentioned above is of course compatible with this derivation as well. In general the data tend to confirm the second derivation, but it is less well supported than the first.

The four means from the passing-of-time scale form a pattern that is clearly compatible with the two derivations just presented. On the average, time seemed to pass more slowly in Hi-Re than in the other three conditions. None of the differences, however, is significant.

Considering the mean percentages of "thinking about food" in the three High Need conditions, we should expect the mean time estimate in Hi-No to fall between the means of Hi-Re and Hi-Ir. Contrary to this expectation it is practically as low as the mean in Hi-Ir. A few Hi-No Ss commented that time seemed rather long but that they decided "it couldn't be that long" and therefore put down a low estimate. In order to learn more about this, four additional Hi-No sessions were conducted in which the Ss were thoroughly questioned about the way they arrived at their time estimates (which again combined to a very low average of 518 sec.). One of these Ss said explicitly that the time seemed long but that she rejected this impression, because she did not believe the *E* would be so unkind as to leave her

alone for a long time with nothing to do. If only a few Hi-No Ss reacted in this way, then the low average time estimate in this condition is no longer a surprise.

Some further evidence supports this interpretation. An S to whom the delay period seems long, but who discounts this impression in favor of a "reasonable" estimate, can be expected to answer with some emphasis after reading the next question that time seemed to pass slowly. According to this we should expect the Hi-No mean on the passing-of-time scale to be rather high, and this is indeed the case (see Table 2). It is significantly higher than any of the other means at $p < .001$.

Interestingly enough, on the index "delay period estimate minus game period estimate" the predicted relationship between Hi-No and Hi-Ir appeared: The Hi-No mean is significantly larger ($p < .04$) than the mean in Hi-Ir. From Table 2 one may infer that the mean game period estimate in Hi-No (576 sec.) is almost as low as the mean delay period estimate and hence much lower than the average game period estimates in all other conditions. We may interpret this finding as an indication that Hi-No Ss did not experience the game period as longer than the delay period, and that remembering their previous estimates they were forced to give equally low estimates for the second period.

On the whole, the data just presented support the hypothesis about time estimation as a function of force quite well. Only one finding does not seem to fall in line with what we should expect: On both time estimation indices, Hi-Ir is lower than Lo-Re and Lo-Ir. If we want to interpret this finding in terms of our hypothesis we have to argue that the force to eat in Hi-Ir Ss was very low at the time of the delay period estimate. This assumption seems to be contradicted by the data from the "How hungry" scale, on which the Hi-Ir mean is significantly higher than the means of the two Low Need conditions. It is debatable, however, whether this scale can be regarded as a good measure of the force to eat in Hi-Ir Ss at the time of the delay period estimate, i.e., immediately after the rating task. After this estimate a series of food-related questions followed, which furnished relevant stimulation. In Hi-Re and Lo-Re, these questions only

continued the stimulation to think about food. In Lo-Ir, we should not expect any marked effects from these questions because need was low in this condition. In Hi-Ir Ss, however, we should predict an increase in the force to eat from the time estimate to the "How hungry" question. Thus we may conclude that the Hi-Ir mean on the "How hungry" scale leads to a considerable overestimation of the force to eat acting on Hi-Ir Ss immediately after the rating.

This conclusion receives good support from the data of the "Desire to taste" scale, which was the first measure of the force to eat in the order of questions presented to S. If we assume that with continuing relevant stimulation from these questions the force to eat increased in Hi-Ir, then we might expect this force to appear still relatively small on the earliest of these indices, and this is the case (see Table 1). Indeed, the mean score is actually lower than in the Lo-Re condition.

Though these considerations *may* explain the finding that, on the average, time estimation in Hi-Ir was lower than in Lo-Re, they cannot, of course, handle the relationship of Hi-Ir and Lo-Ir. Although one is tempted to dismiss the difference between these conditions as nonsignificant and thus in accord with prediction, the lower time estimation in the Hi-Ir condition is clearly disturbing. We do not, at present, wish to offer explanatory suggestions for which there can be no evidence in the immediate experimental data. It seems clear that further investigation of the relationship between force and time perception is required.

EXPERIMENT II

Our experimental design involved one flaw, for the variation of relevance could only be handled by covarying the activity. In relevant conditions, Ss rated recipes; in irrelevant conditions, Ss rated fashion designs. The results for time estimation might therefore be attributable to differing activities rather than to differences in relevance of ideation, and, indeed, there is evidence that possibly supports this argument.

In Table 3 the reader can see that Ss who rated recipes liked the rating task less than Ss who rated fashion designs. The difference between Hi-Re and Hi-Ir is significant at

TABLE 3
MEAN SCORES FOR "LIKING OF THE RATING TASK"
AND MEAN PERCENTAGES OF
"THINKING ABOUT THE GOAL"

Condition	Rated Objects	Liking of the Rating Task ^a	Thinking about the Goal
Experiment I			
Hi-Re	Recipes	3.75	91
Hi-Ir	Fashion designs	4.19	7
Lo-Re	Recipes	3.47	88
Lo-Ir	Fashion designs	4.13	9
Experiment II			
Rel	Fashion designs	4.45	96
Irrel	Recipes	3.62	6

^a The higher the score, the greater the liking.

$p < .03$ and the difference between Lo-Re and Lo-Ir at $p < .001$. Thus we have to realize that the low delay period estimates in Hi-Ir and Lo-Ir are linked both to irrelevant stimulation and high liking of the task, and the high time estimates in Hi-Re and Lo-Re to both relevant stimulation and low liking of the rating period. As a consequence of this, three alternative interpretations of our results can be offered:

1. The association between relevance and liking was a coincidence; liking of the task has no connection with the time estimates.

2. The lower liking scores are indicative of stronger forces acting on Ss in the Relevant Stimulation conditions to leave the rating situation and reach the goal of tasting.

3. Rating recipes was less attractive for our college girls than rating fashion designs, and the lower liking of the recipe rating was the ultimate cause of the higher time estimates in Hi-Re and Lo-Re.

In order to decide among these alternatives, it seemed necessary to design an experiment which, in effect, would be the reverse of the study described, i.e., a study in which rating recipes would be the irrelevant condition and rating fashion designs the relevant condition.

Method

Again college girls, students of introductory psychology and sociology classes at the University of Minnesota served as Ss in the experiment. They were reached by telephone and invited to an experiment on judging of fashions.

At the beginning of each experimental session, which was conducted with one S at a time, E created a highly

desirable goal for *S* in relation to which rating of fashion designs would provide relevant stimulation. He explained that some of the leading fashion houses in Europe and America had decided on a conference to sponsor a study on ladies' tastes and preferences in fashions and that he had been contracted to carry out part of the project in Minneapolis. In a while he would accompany her to a display room in which a representative of the Parisian fashion world would show her a series of cocktail dresses and evening gowns designed by some of those famous fashion houses and so far not yet modeled in public fashion shows. Her task would then be to examine these dresses as long as she wished and to fill in a short questionnaire in which she would be asked to state her preferences among the various models and to make some judgments about details of the designs.

After this introduction, *E* announced a delay due to some necessary preparations, asked the *S* to lend him her watch under the pretext that his was broken and created one of two stimulation conditions much as in Experiment I. With the explanation that he was also interested to learn more about tastes and preferences in general, he either asked *S* to rate fashion designs (Relevant Stimulation) or recipes (Irrelevant Stimulation) during his absence, and left her alone for 13 minutes. Each of these two stimulation conditions, *Rel* and *Irrel*, was run with 17 *Ss*.

After the delay period, *E* returned with a questionnaire in which *S* was asked (a) to estimate the length of this period; and to indicate on graphic rating scales (b) how she liked the rating task, (c) how much she thought about fashions, dresses, etc., and (d) how strong her desire was to start examining the dresses on display, and, finally, to say (e) how many additional hours she would be able to spend on further fashion shows.

When *S* had filled in this questionnaire, *E* announced the end of the experiment, explained it in detail, and made the request not to talk about it, which apparently was again faithfully observed.

Results and Discussion

According to the three alternative interpretations of the time estimation results obtained in the first experiment, three possible outcomes of this second experiment were considered:

1. If recipe raters would again give lower liking scores, but now make lower time estimates than fashion design raters, then we should conclude that in our first experiment the association of relevance of stimulation and liking of the task was a coincidence and that liking of the task had little or no effects on the time estimates.

2. If, on the other hand, recipe raters would now give higher liking scores and lower time estimates than fashion design raters, then we

should conclude that relevance of stimulation was the ultimate cause for the higher time estimates in the first experiment and that the liking scores were indicators of the resultant forces to leave the region of the delay.

3. Finally, if recipe raters would again give lower liking scores and higher time estimates than fashion design raters, then we should say that recipe rating per se was less attractive and thus ultimately responsible for the time estimation results in the first experiment.

Table 3 demonstrates that in the second experiment, rating of recipes distracted from thinking about the promised fashion show just as effectively as the rating of fashion designs had distracted from thinking about food. Yet, again, recipe rating was less liked than rating of fashion designs. The difference between *Rel* and *Irrel* in this respect is significant at $p < .002$. This is clear evidence against the second possible interpretation offered above. Liking of the task was not related to the relevance of the stimulation it provided. To decide between the two remaining alternatives one may look at Table 4.

As predicted by our two hypotheses, the recipe raters now gave lower time estimates than the fashion design raters ($p < .12$). This clearly supports the first interpretation given above that the attractiveness of the rating task had nothing to do with the functional relationship between relevance of ideation and time estimation. Table 4 also shows that the mean number of hours *Ss* felt they could devote to further experiments on fashion rating was somewhat greater in *Rel* than in *Irrel* ($p < .20$), and that the *Rel* *Ss* indicated a greater desire to start examining the real dresses than the *Irrel* *Ss* ($p < .05$). These results furnish additional evidence in favor of our first hypothesis about the effects of relevance of ideation on force.

TABLE 4
MEANS OF THE DATA ON TIME ESTIMATION AND FORCE IN EXPERIMENT II

Condition	Estimate of the Delay Period (sec.)	Desire to Examine Dresses ^a	Additional Hours of Participation
<i>Rel</i>	805	3.65	3.68
<i>Irrel</i>	667	2.92	2.85

^a The higher the score, the greater the desire.

SUMMARY

An exploratory study by Schachter on social isolation suggested two hypotheses:

1. The force acting on a person in a barrier situation to reach a goal G is an increasing function of the person's need for G "times" the relevance of his ideation with respect to G.

2. The greater the magnitude of the force to reach a goal acting on a person in a barrier situation (within limits), the greater will be his estimation of the time spent in the barrier situation.

These hypotheses were tested with female Ss in an experiment on "food tasting," in which deprivation of food, thinking about food, and the desire to eat corresponded to the concepts need, relevance of ideation, and force. The first hypothesis was strongly supported by the experiment and, with the exception of one result, the data also provided good evidence in favor of the second one. The possibility that the time estimations were influenced by the attractiveness of the activities involved in the manipulation of relevance was excluded by the results of a supplementary experiment.

REFERENCES

- BROŽEK, J., GUETZKOW, H., & BALDWIN, M. V. A quantitative study of perception and association in experimental semistarvation. *J. Pers.*, 1951, **19**, 245-264.
- CALVIN, J. S., BICKNELL, E. A., & SPERLING, D. S. Establishment of a conditioned drive based on the hunger drive. *J. comp. physiol. Psychol.*, 1953, **46**, 173-175.
- CLARK, R. A. The projective measurement of experimentally induced sexual motivation. *J. exp. Psychol.*, 1952, **44**, 391-399.
- DANZIGER, K. The operation of an acquired drive in satiated rats. *Quart. J. exp. Psychol.*, 1951, **3**, 119-132.
- FALK, J. L., & BINDRA, D. Judgment of time as a function of serial position and stress. *J. exp. Psychol.*, 1954, **47**, 279-282.
- FILER, R. J., & MEALS, D. W. The effect of motivating conditions on the estimation of time. *J. exp. Psychol.*, 1949, **39**, 327-333.
- GREENBERG, I. The acquisition of a thirst drive. *Dissertation Abstr.*, 1954, **14**, 1465-1466.
- HENRIKSON, E. H. A study of stage fright and the judgment of speaking time. *J. appl. Psychol.*, 1948, **32**, 532-536.
- KLEIN, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska symposium on motivation*. Lincoln: Univer. Nebraska Press, 1954.
- MURPHY, G. *Personality: A biosocial approach to origins and structure*. New York: Harper, 1947.
- POSTMAN, L. The experimental analysis of motivational factors in perception. In M. R. Jones (Ed.), *Current theory and research in motivation*. Lincoln: Univer. Nebraska Press, 1953.
- ROSENZWEIG, S., & KOTH, A. G. The experience of duration as affected by need-tension. *J. exp. Psychol.*, 1933, **16**, 745-775.
- SANFORD, R. N. The effect of abstinence from food upon imaginal processes: A preliminary experiment. *J. Psychol.*, 1936, **2**, 129-136.
- SANFORD, R. N. The effect of abstinence from food upon imaginal processes: A further experiment. *J. Psychol.*, 1937, **3**, 145-159.
- SCHÖNBACH, P. M. Need, relevance of ideation, force and time estimation. Unpublished doctoral dissertation, Univer. of Minnesota, 1956.
- SIEGEL, P. S., & MACDONNELL, M. F. A repetition of the Calvin-Bicknell-Sperling study of conditioned drive. *J. comp. physiol. Psychol.*, 1954, **47**, 250-252.
- WISPÉ, L. G. Physiological need, verbal frequency, and word association. *J. abnorm. soc. Psychol.*, 1954, **49**, 229-234.

Received November 18, 1957.

COGNITIVE CONSEQUENCES OF FORCED COMPLIANCE

LEON FESTINGER AND JAMES M. CARLSMITH¹

Stanford University

WHAT happens to a person's private opinion if he is forced to do or say something contrary to that opinion? Only recently has there been any experimental work related to this question. Two studies reported by Janis and King (1954; 1956) clearly showed that, at least under some conditions, the private opinion changes so as to bring it into closer correspondence with the overt behavior the person was forced to perform. Specifically, they showed that if a person is forced to improvise a speech supporting a point of view with which he disagrees, his private opinion moves toward the position advocated in the speech. The observed opinion change is greater than for persons who only hear the speech or for persons who read a prepared speech with emphasis solely on elocution and manner of delivery. The authors of these two studies explain their results mainly in terms of mental rehearsal and thinking up new arguments. In this way, they propose, the person who is forced to improvise a speech convinces himself. They present some evidence, which is not altogether conclusive, in support of this explanation. We will have more to say concerning this explanation in discussing the results of our experiment.

Kelman (1953) tried to pursue the matter further. He reasoned that if the person is induced to make an overt statement contrary to his private opinion by the offer of some reward, then the greater the reward offered, the greater should be the subsequent opinion change. His data, however, did not support this idea. He found, rather, that a large reward produced less subsequent opinion change than did a smaller reward. Actually, this finding by Kelman is consistent with the theory we will outline below but, for a number of reasons, is

not conclusive. One of the major weaknesses of the data is that not all subjects in the experiment made an overt statement contrary to their private opinion in order to obtain the offered reward. What is more, as one might expect, the percentage of subjects who complied increased as the size of the offered reward increased. Thus, with self-selection of who did and who did not make the required overt statement and with varying percentages of subjects in the different conditions who did make the required statement, no interpretation of the data can be unequivocal.

Recently, Festinger (1957) proposed a theory concerning cognitive dissonance from which come a number of derivations about opinion change following forced compliance. Since these derivations are stated in detail by Festinger (1957, Ch. 4), we will here give only a brief outline of the reasoning.

Let us consider a person who privately holds opinion "X" but has, as a result of pressure brought to bear on him, publicly stated that he believes "not X."

1. This person has two cognitions which, psychologically, do not fit together: one of these is the knowledge that he believes "X," the other the knowledge that he has publicly stated that he believes "not X." If no factors other than his private opinion are considered, it would follow, at least in our culture, that if he believes "X" he would publicly state "X." Hence, his cognition of his private belief is dissonant with his cognition concerning his actual public statement.

2. Similarly, the knowledge that he has said "not X" is consonant with (does fit together with) those cognitive elements corresponding to the reasons, pressures, promises of rewards and/or threats of punishment which induced him to say "not X."

3. In evaluating the total magnitude of dissonance, one must take account of both dissonances and consonances. Let us think of the sum of all the dissonances involving some particular cognition as "D" and the sum of all the consonances as "C." Then we might

¹ The experiment reported here was done as part of a program of research supported by a grant from the National Science Foundation to the senior author. We wish to thank Leonard Hommel, Judson Mills, and Robert Terwilliger for their help in designing and carrying out the experiment. We would also like to acknowledge the help of Ruth Smith and Marilyn M. Miller.

think of the total magnitude of dissonance as being a function of "D" divided by "D" plus "C."

Let us then see what can be said about the total magnitude of dissonance in a person created by the knowledge that he said "not X" and really believes "X." With everything else held constant, this total magnitude of dissonance would decrease as the number and importance of the pressures which induced him to say "not X" increased.

Thus, if the overt behavior was brought about by, say, offers of reward or threats of punishment, the magnitude of dissonance is maximal if these promised rewards or threatened punishments were just barely sufficient to induce the person to say "not X." From this point on, as the promised rewards or threatened punishment become larger, the magnitude of dissonance becomes smaller.

4. One way in which the dissonance can be reduced is for the person to change his private opinion so as to bring it into correspondence with what he has said. One would consequently expect to observe such opinion change after a person has been forced or induced to say something contrary to his private opinion. Furthermore, since the pressure to reduce dissonance will be a function of the magnitude of the dissonance, the observed opinion change should be greatest when the pressure used to elicit the overt behavior is just sufficient to do it.

The present experiment was designed to test this derivation under controlled, laboratory conditions. In the experiment we varied the amount of reward used to force persons to make a statement contrary to their private views. The prediction [from 3 and 4 above] is that the larger the reward given to the subject, the smaller will be the subsequent opinion change.

PROCEDURE

Seventy-one male students in the introductory psychology course at Stanford University were used in the experiment. In this course, students are required to spend a certain number of hours as subjects (Ss) in experiments. They choose among the available experiments by signing their names on a sheet posted on the bulletin board which states the nature of the experiment. The present experiment was listed

as a two-hour experiment dealing with "Measures of Performance."

During the first week of the course, when the requirement of serving in experiments was announced and explained to the students, the instructor also told them about a study that the psychology department was conducting. He explained that, since they were required to serve in experiments, the department was conducting a study to evaluate these experiments in order to be able to improve them in the future. They were told that a sample of students would be interviewed after having served as Ss. They were urged to cooperate in these interviews by being completely frank and honest. The importance of this announcement will become clear shortly. It enabled us to measure the opinions of our Ss in a context not directly connected with our experiment and in which we could reasonably expect frank and honest expressions of opinion.

When the *S* arrived for the experiment on "Measures of Performance" he had to wait for a few minutes in the secretary's office. The experimenter (*E*) then came in, introduced himself to the *S* and, together, they walked into the laboratory room where the *E* said:

This experiment usually takes a little over an hour but, of course, we had to schedule it for two hours. Since we have that extra time, the introductory psychology people asked if they could interview some of our subjects. [Offhand and conversationally.] Did they announce that in class? I gather that they're interviewing some people who have been in experiments. I don't know much about it. Anyhow, they may want to interview you when you're through here.

With no further introduction or explanation the *S* was shown the first task, which involved putting 12 spools onto a tray, emptying the tray, refilling it with spools, and so on. He was told to use one hand and to work at his own speed. He did this for one-half hour. The *E* then removed the tray and spools and placed in front of the *S* a board containing 48 square pegs. His task was to turn each peg a quarter turn clockwise, then another quarter turn, and so on. He was told again to use one hand and to work at his own speed. The *S* worked at this task for another half hour.

While the *S* was working on these tasks, the *E* sat, with a stop watch in his hand, busily making notations on a sheet of paper. He did so in order to make it convincing that this was

what the *E* was interested in and that these tasks, and how the *S* worked on them, was the total experiment. From our point of view the experiment had hardly started. The hour which the *S* spent working on the repetitive, monotonous tasks was intended to provide, for each *S* uniformly, an experience about which he would have a somewhat negative opinion.

After the half hour on the second task was over, the *E* conspicuously set the stop watch back to zero, put it away, pushed his chair back, lit a cigarette, and said:

O.K. Well, that's all we have in the experiment itself. I'd like to explain what this has been all about so you'll have some idea of why you were doing this. [E pauses.] Well, the way the experiment is set up is this. There are actually two groups in the experiment. In one, the group you were in, we bring the subject in and give him essentially no introduction to the experiment. That is, all we tell him is what he needs to know in order to do the tasks, and he has no idea of what the experiment is all about, or what it's going to be like, or anything like that. But in the other group, we have a student that we've hired that works for us regularly, and what I do is take him into the next room where the subject is waiting—the same room you were waiting in before—and I introduce him as if he had just finished being a subject in the experiment. That is, I say: "This is so-and-so, who's just finished the experiment, and I've asked him to tell you a little of what it's about before you start." The fellow who works for us then, in conversation with the next subject, makes these points: [The *E* then produced a sheet headed "For Group B" which had written on it: It was very enjoyable, I had a lot of fun, I enjoyed myself, it was very interesting, it was intriguing, it was exciting. The *E* showed this to the *S* and then proceeded with his false explanation of the purpose of the experiment.] Now, of course, we have this student do this, because if the experimenter does it, it doesn't look as realistic, and what we're interested in doing is comparing how these two groups do on the experiment—the one with this previous expectation about the experiment, and the other, like yourself, with essentially none.

Up to this point the procedure was identical for *Ss* in all conditions. From this point on they diverged somewhat. Three conditions were run, Control, One Dollar, and Twenty Dollars, as follows:

Control Condition

The *E* continued:

Is that fairly clear? [Pause.] Look, that fellow [looks at watch] I was telling you about from the introductory psychology class said he would get here a couple of minutes from now. Would you mind waiting to see if he wants to talk to you? Fine. Why don't we go into

the other room to wait? [The *E* left the *S* in the secretary's office for four minutes. He then returned and said:] O.K. Let's check and see if he does want to talk to you.

One and Twenty Dollar Conditions

The *E* continued:

Is that fairly clear how it is set up and what we're trying to do? [Pause.] Now, I also have a sort of strange thing to ask you. The thing is this. [Long pause, some confusion and uncertainty in the following, with a degree of embarrassment on the part of the *E*. The manner of the *E* contrasted strongly with the preceding unhesitant and assured false explanation of the experiment. The point was to make it seem to the *S* that this was the first time the *E* had done this and that he felt unsure of himself.] The fellow who normally does this for us couldn't do it today—he just phoned in, and something or other came up for him—so we've been looking around for someone that we could hire to do it for us. You see, we've got another subject waiting [looks at watch] who is supposed to be in that other condition. Now Professor ———, who is in charge of this experiment, suggested that perhaps we could take a chance on your doing it for us. I'll tell you what we had in mind: the thing is, if you could do it for us now, then of course you would know how to do it, and if something like this should ever come up again, that is, the regular fellow couldn't make it, and we had a subject scheduled, it would be very reassuring to us to know that we had somebody else we could call on who knew how to do it. So, if you would be willing to do this for us, we'd like to hire you to do it now and then be on call in the future, if something like this should ever happen again. We can pay you a dollar (twenty dollars) for doing this for us, that is, for doing it now and then being on call. Do you think you could do that for us?

If the *S* hesitated, the *E* said things like, "It will only take a few minutes," "The regular person is pretty reliable; this is the first time he has missed," or "If we needed you we could phone you a day or two in advance; if you couldn't make it, of course, we wouldn't expect you to come." After the *S* agreed to do it, the *E* gave him the previously mentioned sheet of paper headed "For Group B" and asked him to read it through again. The *E* then paid the *S* one dollar (twenty dollars), made out a hand-written receipt form, and asked the *S* to sign it. He then said:

O.K., the way we'll do it is this. As I said, the next subject should be here by now. I think the next one is a girl. I'll take you into the next room and introduce you to her, saying that you've just finished the experiment and that we've asked you to tell her a little about it. And what we want you to do is just sit down and get into a conversation with her and try to get

across the points on that sheet of paper. I'll leave you alone and come back after a couple of minutes. O.K.?

The *E* then took the *S* into the secretary's office where he had previously waited and where the next *S* was waiting. (The secretary had left the office.) He introduced the girl and the *S* to one another saying that the *S* had just finished the experiment and would tell her something about it. He then left saying he would return in a couple of minutes. The girl, an undergraduate hired for this role, said little until the *S* made some positive remarks about the experiment and then said that she was surprised because a friend of hers had taken the experiment the week before and had told her that it was boring and that she ought to try to get out of it. Most *S*s responded by saying something like "Oh, no, it's really very interesting. I'm sure you'll enjoy it." The girl, after this listened quietly, accepting and agreeing to everything the *S* told her. The discussion between the *S* and the girl was recorded on a hidden tape recorder.

After two minutes the *E* returned, asked the girl to go into the experimental room, thanked the *S* for talking to the girl, wrote down his phone number to continue the fiction that we might call on him again in the future and then said: "Look, could we check and see if that fellow from introductory psychology wants to talk to you?"

From this point on, the procedure for all three conditions was once more identical. As the *E* and the *S* started to walk to the office where the interviewer was, the *E* said: "Thanks very much for working on those tasks for us. I hope you did enjoy it. Most of our subjects tell us afterward that they found it quite interesting. You get a chance to see how you react to the tasks and so forth." This short persuasive communication was made in all conditions in exactly the same way. The reason for doing it, theoretically, was to make it easier for anyone who wanted to persuade himself that the tasks had been, indeed, enjoyable.

When they arrived at the interviewer's office, the *E* asked the interviewer whether or not he wanted to talk to the *S*. The interviewer said yes, the *E* shook hands with the *S*, said good-bye, and left. The interviewer, of course, was always kept in complete ignorance of which condition the *S* was in. The interview

consisted of four questions, on each of which the *S* was first encouraged to talk about the matter and was then asked to rate his opinion or reaction on an 11-point scale. The questions are as follows:

1. Were the tasks interesting and enjoyable? In what way? In what way were they not? Would you rate how you feel about them on a scale from -5 to +5 where -5 means they were extremely dull and boring, +5 means they were extremely interesting and enjoyable, and zero means they were neutral, neither interesting nor uninteresting.

2. Did the experiment give you an opportunity to learn about your own ability to perform these tasks? In what way? In what way not? Would you rate how you feel about this on a scale from 0 to 10 where 0 means you learned nothing and 10 means you learned a great deal.

3. From what you know about the experiment and the tasks involved in it, would you say the experiment was measuring anything important? That is, do you think the results may have scientific value? In what way? In what way not? Would you rate your opinion on this matter on a scale from 0 to 10 where 0 means the results have no scientific value or importance and 10 means they have a great deal of value and importance.

4. Would you have any desire to participate in another similar experiment? Why? Why not? Would you rate your desire to participate in a similar experiment again on a scale from -5 to +5, where -5 means you would definitely dislike to participate, +5 means you would definitely like to participate, and 0 means you have no particular feeling about it one way or the other.

As may be seen, the questions varied in how directly relevant they were to what the *S* had told the girl. This point will be discussed further in connection with the results.

At the close of the interview the *S* was asked what he thought the experiment was about and following this, was asked directly whether or not he was suspicious of anything and, if so, what he was suspicious of. When the interview was over, the interviewer brought the *S* back to the experimental room where the *E* was waiting together with the girl who had posed as the waiting *S*. (In the control condition, of course, the girl was not there.) The true purpose of the experiment was then explained to the *S* in detail, and the reasons for each of the various steps in the experiment were explained carefully in relation to the true purpose. All experimental *S*s in both One Dollar and Twenty Dollar conditions were asked, after this explanation, to return the money they had

been given. All Ss, without exception, were quite willing to return the money.

The data from 11 of the 71 Ss in the experiment had to be discarded for the following reasons:

1. Five Ss (three in the One Dollar and two in the Twenty Dollar condition) indicated in the interview that they were suspicious about having been paid to tell the girl the experiment was fun and suspected that that was the real purpose of the experiment.
2. Two Ss (both in the One Dollar condition) told the girl that they had been hired, that the experiment was really boring but they were supposed to say it was fun.
3. Three Ss (one in the One Dollar and two in the Twenty Dollar condition) refused to take the money and refused to be hired.
4. One S (in the One Dollar condition), immediately after having talked to the girl, demanded her phone number saying he would call her and explain things, and also told the E he wanted to wait until she was finished so he could tell her about it.

These 11 Ss were, of course, run through the total experiment anyhow and the experiment was explained to them afterwards. Their data, however, are not included in the analysis.

Summary of Design

There remain, for analysis, 20 Ss in each of the three conditions. Let us review these briefly: 1. *Control condition*. These Ss were treated identically in all respects to the Ss in the experimental conditions, except that they were never asked to, and never did, tell the waiting girl that the experimental tasks were enjoyable and lots of fun. 2. *One Dollar condition*. These Ss were hired for one dollar to tell a waiting S that tasks, which were really rather dull and boring, were interesting, enjoyable, and lots of fun. 3. *Twenty Dollar condition*. These Ss were hired for twenty dollars to do the same thing.

RESULTS

The major results of the experiment are summarized in Table 1 which lists, separately for each of the three experimental conditions, the average rating which the Ss gave at the end of each question on the interview. We will discuss each of the questions on the interview separately, because they were intended to measure different things. One other point before we proceed to examine the data. In all the comparisons, the Control condition should be

TABLE 1
AVERAGE RATINGS ON INTERVIEW QUESTIONS FOR EACH CONDITION

Question on Interview	Experimental Condition		
	Control (N = 20)	One Dollar (N = 20)	Twenty Dollars (N = 20)
How enjoyable tasks were (rated from -5 to +5)	-.45	+1.35	-.05
How much they learned (rated from 0 to 10)	3.08	2.80	3.15
Scientific importance (rated from 0 to 10)	5.60	6.45	5.18
Participate in similar exp. (rated from -5 to +5)	-.62	+1.20	-.25

regarded as a baseline from which to evaluate the results in the other two conditions. The Control condition gives us, essentially, the reactions of Ss to the tasks and their opinions about the experiment as falsely explained to them, without the experimental introduction of dissonance. The data from the other conditions may be viewed, in a sense, as changes from this baseline.

How Enjoyable the Tasks Were

The average ratings on this question, presented in the first row of figures in Table 1, are the results most important to the experiment. These results are the ones most directly relevant to the specific dissonance which was experimentally created. It will be recalled that the tasks were purposely arranged to be rather boring and monotonous. And, indeed, in the Control condition the average rating was $-.45$, somewhat on the negative side of the neutral point.

In the other two conditions, however, the Ss told someone that these tasks were interesting and enjoyable. The resulting dissonance could, of course, most directly be reduced by persuading themselves that the tasks were, indeed, interesting and enjoyable. In the One Dollar condition, since the magnitude of dissonance was high, the pressure to reduce this dissonance would also be high. In this condition, the average rating was $+1.35$, considerably on the positive side and significantly different from the Control condition at the .02 level² ($t = 2.48$).

² All statistical tests referred to in this paper are two-tailed.

In the Twenty Dollar condition, where less dissonance was created experimentally because of the greater importance of the consonant relations, there is correspondingly less evidence of dissonance reduction. The average rating in this condition is only $-.05$, slightly and not significantly higher than the Control condition. The difference between the One Dollar and Twenty Dollar conditions is significant at the $.03$ level ($t = 2.22$). In short, when an S was induced, by offer of reward, to say something contrary to his private opinion, this private opinion tended to change so as to correspond more closely with what he had said. The greater the reward offered (beyond what was necessary to elicit the behavior) the smaller was the effect.

Desire to Participate in a Similar Experiment

The results from this question are shown in the last row of Table 1. This question is less directly related to the dissonance that was experimentally created for the S s. Certainly, the more interesting and enjoyable they felt the tasks were, the greater would be their desire to participate in a similar experiment. But other factors would enter also. Hence, one would expect the results on this question to be very similar to the results on "how enjoyable the tasks were" but weaker. Actually, the result, as may be seen in the table, are in exactly the same direction, and the magnitude of the mean differences is fully as large as on the first question. The variability is greater, however, and the differences do not yield high levels of statistical significance. The difference between the One Dollar condition ($+1.20$) and the Control condition ($-.62$) is significant at the $.08$ level ($t = 1.78$). The difference between the One Dollar condition and the Twenty Dollar condition ($-.25$) reaches only the $.15$ level of significance ($t = 1.46$).

The Scientific Importance of the Experiment

This question was included because there was a chance that differences might emerge. There are, after all, other ways in which the experimentally created dissonance could be reduced. For example, one way would be for the S to magnify for himself the value of the reward he obtained. This, however, was un-

likely in this experiment because money was used for the reward and it is undoubtedly difficult to convince oneself that one dollar is more than it really is. There is another possible way, however. The S s were given a very good reason, in addition to being paid, for saying what they did to the waiting girl. The S s were told it was necessary for the experiment. The dissonance could, consequently, be reduced by magnifying the importance of this cognition. The more scientifically important they considered the experiment to be, the less was the total magnitude of dissonance. It is possible, then, that the results on this question, shown in the third row of figures in Table 1, might reflect dissonance reduction.

The results are weakly in line with what one would expect if the dissonance were somewhat reduced in this manner. The One Dollar condition is higher than the other two. The difference between the One and Twenty Dollar conditions reaches the $.08$ level of significance on a two-tailed test ($t = 1.79$). The difference between the One Dollar and Control conditions is not impressive at all ($t = 1.21$). The result that the Twenty Dollar condition is actually lower than the Control condition is undoubtedly a matter of chance ($t = 0.58$).

How Much They Learned From the Experiment

The results on this question are shown in the second row of figures in Table 1. The question was included because, as far as we could see, it had nothing to do with the dissonance that was experimentally created and could not be used for dissonance reduction. One would then expect no differences at all among the three conditions. We felt it was important to show that the effect was not a completely general one but was specific to the content of the dissonance which was created. As can be readily seen in Table 1, there are only negligible differences among conditions. The highest t value for any of these differences is only 0.48 .

DISCUSSION OF A POSSIBLE ALTERNATIVE EXPLANATION

We mentioned in the introduction that Janis and King (1954; 1956) in explaining their findings, proposed an explanation in terms of the self-convincing effect of mental rehearsal

and thinking up new arguments by the person who had to improvise a speech. Kelman (1953), in the previously mentioned study, in attempting to explain the unexpected finding that the persons who complied in the moderate reward condition changed their opinion more than in the high reward condition, also proposed the same kind of explanation. If the results of our experiment are to be taken as strong corroboration of the theory of cognitive dissonance, this possible alternative explanation must be dealt with.

Specifically, as applied to our results, this alternative explanation would maintain that perhaps, for some reason, the *Ss* in the One Dollar condition worked harder at telling the waiting girl that the tasks were fun and enjoyable. That is, in the One Dollar condition they may have rehearsed it more mentally, thought up more ways of saying it, may have said it more convincingly, and so on. Why this might have been the case is, of course, not immediately apparent. One might expect that, in the Twenty Dollar condition, having been paid more, they would try to do a better job of it than in the One Dollar condition. But nevertheless, the possibility exists that the *Ss* in the One Dollar condition may have improvised more.

Because of the desirability of investigating this possible alternative explanation, we recorded on a tape recorder the conversation between each *S* and the girl. These recordings were transcribed and then rated, by two independent raters, on five dimensions. The ratings were, of course done in ignorance of which condition each *S* was in. The reliabilities of these ratings, that is, the correlations from the two independent raters, ranged from .61 to .88, with an average reliability of .71. The five ratings were:

1. The content of what the *S* said *before* the girl made the remark that her friend told her it was boring. The stronger the *S*'s positive statements about the tasks, and the more ways in which he said they were interesting and enjoyable, the higher the rating.

2. The content of what the *S* said *after* the girl made the above-mentioned remark. This was rated in the same way as for the content before the remark.

3. A similar rating of the over-all content of what the *S* said.

4. A rating of how persuasive and convincing the *S* was in what he said and the way in which he said it.

5. A rating of the amount of time in the discussion that the *S* spent discussing the tasks as opposed to going off into irrelevant things.

The mean ratings for the One Dollar and Twenty Dollar conditions, averaging the ratings of the two independent raters, are presented in Table 2. It is clear from examining the table that, in all cases, the Twenty Dollar condition is slightly higher. The differences are small, however, and only on the rating of "amount of time" does the difference between the two conditions even approach significance. We are certainly justified in concluding that the *Ss* in the One Dollar condition did not improvise more nor act more convincingly. Hence, the alternative explanation discussed above cannot account for the findings.

SUMMARY

Recently, Festinger (1957) has proposed a theory concerning cognitive dissonance. Two derivations from this theory are tested here. These are:

1. If a person is induced to do or say something which is contrary to his private opinion, there will be a tendency for him to change his opinion so as to bring it into correspondence with what he has done or said.

2. The larger the pressure used to elicit the

TABLE 2
AVERAGE RATINGS OF DISCUSSION BETWEEN SUBJECT
AND GIRL

Dimension Rated	Condition		
	One Dollar	Twenty Dollars	Value of <i>t</i>
Content before remark by girl (rated from 0 to 5)	2.26	2.62	1.08
Content after remark by girl (rated from 0 to 5)	1.63	1.75	0.11
Over-all content (rated from 0 to 5)	1.89	2.19	1.08
Persuasiveness and conviction (rated from 0 to 10)	4.79	5.50	0.99
Time spent on topic (rated from 0 to 10)	6.74	8.19	1.80

overt behavior (beyond the minimum needed to elicit it) the weaker will be the above-mentioned tendency.

A laboratory experiment was designed to test these derivations. Subjects were subjected to a boring experience and then paid to tell someone that the experience had been interesting and enjoyable. The amount of money paid the subject was varied. The private opinions of the subjects concerning the experiences were then determined.

The results strongly corroborate the theory that was tested.

REFERENCES

- FESTINGER, L. *A theory of cognitive dissonance*. Evanston, Ill: Row Peterson, 1957.
- JANIS, I. L., & KING, B. T. The influence of role-playing on opinion change. *J. abnorm. soc. Psychol.*, 1954, **49**, 211-218.
- KELMAN, H. Attitude change as a function of response restriction. *Hum. Relat.*, 1953, **6**, 185-214.
- KING, B. T., & JANIS, I. L. Comparison of the effectiveness of improvised versus non-improvised role-playing in producing opinion changes. *Hum. Relat.*, 1956, **9**, 177-186.

Received November 18, 1957.

THE STABILITY OF THE SELF-CONCEPT IN ADOLESCENCE¹

MARY ENGEL²

Michael Reese Hospital, Chicago, Illinois

RECENT theory and research point to the importance of the self-concept in understanding and predicting constancies as well as changes in behavior (Brownfain, 1952; Rogers & Dymond, 1954; Taylor, 1955). It is generally believed that an individual's concept of himself achieves a rather high degree of organization during the course of development and comes to resist change once self-differentiation and self-definition have taken place (Lecky, 1945). As yet it is not known by what age the process of self-definition reaches stability. While we know that the concept of self remains relatively stable, even over extended periods of time, in young adults (Taylor, 1955), and while there are a number of theoretical and partially supported statements in the literature about the storms and stresses of certain aspects of adolescent development (Hall, 1904; Kuhlen, 1948), the fate of the self-concept in adolescence is still a matter for speculation. The studies that examine individual differences in the self-concepts of adolescents from a number of vantage points and in several settings (Balester, 1955; Blodgett, 1953; De Lisle, 1953), represent an inroad into the area of self-concept development. However, it is the longitudinal approach that is most appropriate when seeking answers to questions of development.

The primary purpose of the present study was to investigate the stability of the self-concept in adolescence over a two-year period. It was also its purpose to examine the relationship between whatever stability is found and the quality of the self-concept. The interrelationship between self-concept stability, quality

of the self-concept, and several indices of adjustment was also examined.

METHOD

The data were obtained by testing and retesting 172 public school students, 104 of whom were in the eighth grade and 68 of whom were in the tenth grade at the time of the first testing. The same students served as subjects in 1954 and in 1956.³ Table 1 presents the grade and sex distribution of Ss in the two-year study. An analysis of the fathers' occupations revealed that the Ss were mostly of lower-middle and middle-class background.

The hypotheses were formulated in 1954. Their testing required the use of the following measures:

1. Self concept Q sort, paper and pencil form, consisting of items relevant to adolescent concerns.⁴
2. Verbal Subscale of the Differential Aptitude Test, as an estimate of intelligence.
3. Scales *D*, *Pd*, and *K* of the MMPI, as measures of adjustment and "defensiveness."
4. Peer Rating Scale, as a sociometric assessment of adjustment, based on the model provided by Tudenhham (1952).
5. Teachers' Forced Choice Test as another independent measure of adjustment, developed by Ullman (1952).

The set of Q-sort items for the assessment of the self-concept in adolescents was developed along lines largely in conformity with the principles put forth by Stephenson (1935). Briefly, a large pool of items was gathered covering areas of adolescent self-concern as empirically defined by Jersild (1952). The pooled judgments of psychologists, nonprofessional adults, and adolescents were used to reduce and refine the original set, 100 Q-sort items being retained. Judges could agree with demonstrable certainty that these items represent either positively or negatively toned self-referent attitudes. Examples are: "I can take criticism without resentment." "I see little about myself that's outstanding."

In responding, Ss had to distribute the 50 positively and 50 negatively toned items into 11 categories,

³ There were 243 Ss in 1954; the discrepancy between the 1954 and 1956 *N* can be accounted for by attrition during the two-year period. Detailed analysis of data from the attrition group will be presented elsewhere. Whereas the over-all *N* of the longitudinal sample was 172, an *N* of approximately 149 was available for the testing of certain hypotheses, due to the absence of some Ss on some of the testing days in 1956.

⁴ Copyright applied for. A complete list of Q-sort items is included in University Microfilms Publication: Mic 57-2914. Send \$2.25 to University Microfilms, 313 No. First Street, Ann Arbor, Michigan.

¹ Based upon a dissertation submitted in partial fulfillment of the requirements for the Ph.D. degree, George Peabody College. The writer wishes to express her gratitude to Nicholas Hobbs and Julius Seeman for their guidance. She is also indebted to the Vanderbilt-Peabody Self Concept Research Group for helpful suggestions and comments.

² Written while USPHS postdoctoral clinical research fellow at The Menninger Foundation (MF-6502-C).

TABLE 1
SEX DISTRIBUTION OF SUBJECTS

	8th-10th Grade	10th-12th Grade
Boys.....	48	28
Girls.....	56	40
Total.....	104	68
Grand Total.....	172	

ranging from "most like me" to "least like me." The frequency distribution of items was as follows:

Number of items	4	7	9	11	12	14	12	11	9	7	4
Category	1	2	3	4	5	6	7	8	9	10	11

Paper and pencil administration incurs some errors of measurement, probably not pertaining to item-sampling, that are not involved when the usual card sorting procedure is used. The test-retest reliability of the instrument was .68 over a ten-day period with an N of 23 (tenth grade students). This reliability figure was obtained by correlating the values assigned to each item, by each S , on two occasions and represents the mean of 23 correlations (z transformations were used in computing the mean r). It is slightly lower than similar statistics obtained by others, using the card sort (Taylor, 1955).

The maximum positiveness score that can be obtained on the Q sort used in this study is 600. A score of this magnitude would result from placing every one of the 50 positive items in the "most like me" end of the continuum. Placing an equal number of positive and negative items on the upper and the lower end of the continuum would give rise to a score of 300, the point of ambivalence. Customarily, the negative self-concept is defined as a positiveness score falling below the point of ambivalence, whereas the positive self-concept is usually defined as a positiveness score above the point of ambivalence.

In responding to the Peer Rating Scale, each member of a class writes down one to three names of others who seem to suit some brief behavioral descriptions, for example: "Who is the good sport, the person who always plays fair?" "Who gets mad easily and loses his or her temper often?" These descriptions can be roughly ordered along an adjustment-maladjustment continuum. Each S receives a score that reflects the extent to which his peers see him as well functioning in the school situation. The reliability of the Peer Rating Scale was established by test-retest of 22 S s (ninth graders) over a one-week interval. The resulting value of .96 indicates that the adjustment scores derived from ratings of any one subject by the group as a whole are highly reliable.

RESULTS AND DISCUSSION

Stability of Self-Concept

Analysis of the data obtained in 1954 from S s who subsequently dropped out of school indicates that certain important personality differences may have existed between those

TABLE 2
THE STABILITY OF THE SELF-CONCEPT OVER A TWO-YEAR PERIOD ITEM-BY-ITEM CORRELATIONS OF Q SORTS IN 1954 AND 1956

Group	N	Mean z	z	r Cor- responding to Mean z Scores
Girls				
8th-10th grade	45	.6107	.2059	.54
10th-12th grade	37	.6794	.2204	.59
Boys				
8th-10th grade	44	.4775	.2636	.45
10th-12th grade	23	.6004	.2222	.54
Mean		.5919		.53

who left and those who remained in the school.⁵ Because of the strong possibility of selective attrition, caution is indicated when generalizing from the results of the present study.

It was expected that the S s would form three groups with regard to the self-concept: those maintaining positive self-regarding attitudes, those with negative self-regarding attitudes, and those with defensively positive self-concepts. Hypotheses were formulated on the basis of this expectation. All predictions were made in 1954 and were tested in 1956.

It was hypothesized that the self-concept of adolescents would be relatively stable over the two-year period. This hypothesis implies that the stability, internal organization, and crystallization of the self-concept is achieved earlier in development. Stability was defined by relatively high correlations between self-concept Q sorts in 1954 and 1956. Relevant data are presented in Table 2. The over-all mean correlation of .53, for all S s, indicates the extent of stability of the self-concept of adolescents over a two-year period, between grades eight and ten, and ten and twelve. Corrected for attenuation, the over-all mean correlation between the self-concept in 1954 and 1956 is .78.

It was also predicted that the self-concept of S s with a positive attitude toward themselves in 1954 would be significantly more stable over the two-year period than the self-concept of S s with a negative or defensive-positive self-concept. Results bearing on this prediction are presented in Table 3, in which the negative self-concept is defined by scores

⁵ Analysis of personality differences between S s in the longitudinal sample and the attrition group will be presented elsewhere.

falling in the lower 20% of the distribution of self-concept scores, and positive self-concept by scores in the upper 80%. Where the self-concept was positive, and *S* also obtained a *K* score greater than 17 (measure of "defensiveness" derived from the MMPI), *S* was classified as manifesting a defensive-positive self-concept.

To test the hypothesis, correlations between *Q* sorts were converted into *z* scores as measures of stability. An over-all *F* test of differences in stability between self-concept groups resulted in an *F* ratio of 28.12, greatly exceeding the ratio of 5.30 needed for significance at $p = .05$. Individual *t* tests between groups support the conclusions that (a) *Ss* whose self-concept was positive in 1954 were significantly more stable over the two-year period than *Ss* who had negative self-concepts in 1954; (b) *Ss* whose self-concept was defensive-positive in 1954 were significantly more stable than those who had negative self-concepts; (c) *Ss* whose self-concept was positive in 1954 did not differ significantly in stability from those whose self-concept was defensive-positive in 1954.

The prediction that older and younger *Ss* would not differ significantly in stability of self-concept over the two-year period was supported. Age group differences in magnitude of *Q* sort correlations (self-concept stability) resulted in a *t* ratio of .60.

It was also expected that stability of the self-concept would be statistically unrelated

TABLE 3
COMPARISON OF SELF-CONCEPT STABILITY BETWEEN
POSITIVE, NEGATIVE, AND DEFENSIVE-
POSITIVE SELF-CONCEPT GROUPS
OVER THE TWO-YEAR PERIOD

Groups	<i>N</i> ^a	Per Cent of Total <i>N</i> (172)	Mean Stabil- ity ^b	<i>s</i>	<i>t</i>
Positive Self- concept	106	62	.6928	.2060	7.61*
Negative Self- concept	34	20	.3383	.1977	6.99*
Defensive-Positive Self-concept	32	18	.6379	.2138	
Defensive-Positive Self-concept and Positive Self- concept	—	—	—		1.30

* Significant beyond the .05 level.

^a Classification on basis of 1954 data.

^b Based on 1954 and 1956 data, total *N* for this column 149.

TABLE 4
POSITIVENESS OF THE SELF-CONCEPT IN
1954 AND IN 1956

	1954			1956		
	<i>N</i>	<i>M</i>	<i>s</i>	<i>N</i>	<i>M</i>	<i>s</i>
Girls						
8th-10th grade.....	56	359.98	32.01	45	362.76	29.41
10th-12th grade.....	40	358.40	36.82	37	365.59	38.44
Boys						
8th-10th grade.....	48	351.29	34.68	45	352.25	37.15
10th-12th grade.....	27	360.81	23.03	24	369.75	25.72

to intelligence. Testing this prediction required correlating verbal intelligence scores (DAT) with self-concept stability scores. Correlations were nonsignificant, lending support to the hypothesis, except in the case of the tenth-twelfth grade girls, where an *r* of .36 was found between these two variables, which, with an *N* of 35, was significant beyond the .05 level.

On the assumption that cultural ambiguities concerning sex roles should be more likely to affect girls than boys, it was hypothesized that the self-concept of boys would be significantly more stable, over the two-year period; than that of girls. This hypothesis was not upheld. The comparison of the mean stability between boys and girls resulted in a *t* ratio of .76.

In comparing the mean positiveness scores of the *Ss* in 1954 and in 1956 (Table 4) we found an unpredicted increase in mean positiveness. With the sexes combined, both grades shifted in a positive direction, the mean shift being significant beyond the .05 level in case of the older group ($t = -2.44$).

Stability of Self-Concept and Adjustment

The relationship between the stability of the self-concept and three measures of adjustment (teacher ratings, peer ratings, and MMPI measures) was explored through the following prediction: *Ss* who persist in a positive self-regarding attitude should be better adjusted, in terms of the MMPI, teacher ratings, and peer ratings, than those who persist in negative or defensive-positive self-concepts. Table 5 summarizes the "fate" of the quality of the self-concept for all *Ss* over the two-year period.⁶ More detailed analysis revealed that

⁶ The method of categorization used is too detailed for presentation here but is described in detail elsewhere (Engel, 1956).

most of the shift in self-concept quality occurred in the negative self-concept group. Ss who were classified as having negative self-concepts in 1954 more closely approached the mean by 1956. Such shift could be attributed to regression, except that no such shifting toward the mean took place in the case of Ss originally giving evidence of a positive self-concept.

In applying analyses of variance to adjustment indices between groups, 1956 adjustment measures were used. Table 6 shows that *F* ratios on MMPI scores were significant, whereas *F* ratios based on other adjustment measures were not.

Differences in MMPI measures were further examined by individual *t* tests applied to the column means. MMPI adjustment measures showed the group maintaining negative self-concepts to be significantly less well adjusted

(scoring higher on *D* and *Pd*) in 1956 than others, partially upholding the hypothesis.

Concomitance of Change in Self-Concept and in Adjustment

It was predicted that a change in self-concept in the positive direction would be related to improved adjustment, and a change in self-concept in the negative direction would be related to impaired adjustment. For the purpose of testing this hypothesis Ss were regrouped and considered either "positive shifters" or "negative shifters" depending on a change of 20 points away from their original positive self-concept score either in the positive or negative direction. Only Ss on whom full sets of adjustment scores were available were included in this analysis. Adjustment scores for 1956 were subtracted from 1954 adjustment scores and *t* tests were applied to the mean difference scores. Table 7 presents the results bearing on this hypothesis, and supports the conclusion that "negative shifters" obtained significantly higher *Pd* and *D* scores in 1956 as predicted; however, "positive shifters" be-

TABLE 5
THE DISTRIBUTION OF ALL SUBJECTS IN THE
LONGITUDINAL SAMPLE WITH REGARD TO
THE CHANGES AND CONSTANCIES OF
THE SELF-CONCEPT AS SEEN IN 1956

Changes and Constancies of the Self-Concept Between 1954 and 1956	Number	Per Cent of Total Number Sub- jects
Maintained positive self-concept	76	44
Maintained negative self-concept	14	8
Maintained defensive-positive self-concept	11	6
Was defensive-positive in 1954 but did not maintain either defensiveness or positiveness of self-concept	16	9
Was positive but shifted to negative self-concept by more than 20 points	15	9
Was negative but shifted to positive by more than 20 points	17	10
Absent on more than one testing session in 1956 (unclassified)	23	13
Total	172	99

TABLE 6
COLUMN MEANS AND *F* RATIOS FOR THREE
SELF-CONCEPT GROUPS ON MEASURES OF
ADJUSTMENT (1956 MEASURES USED)

Measures	Maintaining Positive Self- Concept		Maintaining Negative Self- Concept		Maintaining Positive Defensive Self-Concept		<i>F</i>
	<i>N</i>	<i>M</i>	<i>N</i>	<i>M</i>	<i>N</i>	<i>M</i>	
<i>Pd</i>	73	13.51	12	20.17	11	13.64	15.27*
<i>D</i>	73	14.90	12	22.25	11	15.18	21.20*
Peer rating	71	232.62	12	175.17	10	268.70	2.77
Teacher rating	72	22.00	14	18.86	10	23.00	2.18

* Significant beyond the .05 level.

TABLE 7
CHANGES IN ADJUSTMENT MEASURES CONCOMITANT WITH SHIFTS IN SELF-CONCEPT
(BASED ON DIFFERENCE SCORES; 1956 SCORES SUBTRACTED FROM 1954 MEASURES)

Adjustment Measures	"Positive Shifters"				"Negative Shifters"			
	<i>N</i>	<i>M</i>	<i>s</i>	<i>t</i>	<i>N</i>	<i>M</i>	<i>s</i>	<i>t</i>
<i>Pd</i> (MMPI)	30	-.47	3.83	.67	13	-3.15	3.53	-3.22**
<i>D</i> (MMPI)	35	-.60	4.89	.73	15	-3.80	4.75	-3.10**
<i>K</i> (MMPI)	35	-2.94	4.41	-3.95*	22	.23	4.85	.22
Teacher rating	40	.83	7.33	.02	22	.23	4.85	.23
Peer rating	37	-25.97	66.55	-2.37**	17	15.53	93.24	-.69

* Significant beyond the .05 level and in the direction opposite from the predicted one.

** Significant beyond the .05 level and in the predicted direction.

came more "defensive" in that they obtained significantly higher *K* scores in 1956 than in 1954; "positive shifters" were seen as significantly more well adjusted by their peers in 1956 than in 1954. Changes in teacher ratings did not differentiate between groups. Thus, this final hypothesis was only partially confirmed.

It should be borne in mind that this study explored mainly one aspect of the self-concept, the conscious self-concept. It may well be that in spite of the consistencies found in adolescents over a two-year period, considerable changes took place in aspects of the concept of self that are less readily admissible into awareness. The exploration of self-concept consistency and its concomitants on a deeper level of personality would require a clinical approach which was precluded by the use of a fairly large number of *Ss* in the present study.

SUMMARY

A study of the stability of the self-concept over two years in adolescence resulted in the following conclusions:

1. Relative stability of the self-concept was demonstrated by an over-all item-by-item correlation of .53 between *Q* sorts obtained in 1954 and in 1956, with an instrument of which the ten-day test-retest reliability was .68.
2. Subjects whose self-concept was negative at the first testing were significantly less stable in self-concept than subjects whose self-concept was positive.
3. Subjects who persisted in a negative self-concept over the two-year period gave evidence of significantly more maladjustment than subjects who persisted in a positive self-concept, when maladjustment is measured by high scores of scales *Pd* and *D* of the MMPI.
4. Subjects who showed less regard for themselves on the *Q* sort on retest, also shifted toward significantly more maladjustment on scales *Pd* and *D* of the MMPI.
5. Subjects who showed more regard for themselves on the *Q* sort on retest, also shifted

toward significantly more adjustment on peer ratings.

6. The positive self-concept scores increased significantly between the two testings for the tenth-twelfth grade subjects, an increase which could not be attributed entirely to the effect of regression.

REFERENCES

- BALESTER, R. S. The self-concept and juvenile delinquency. Unpublished doctoral dissertation, Vanderbilt Univer., 1955.
- BLODGETT, HARRIETT E. An experimental approach to the measurement of self-evaluation among adolescent girls. Unpublished doctoral dissertation, Univer. of Minnesota, 1953.
- BROWNFAIN, J. J. Stability of self-concept as a dimension of personality. *J. abnorm. soc. Psychol.*, 1952, **47**, 597-606.
- DELISLE, F. H. A study of the relationship of the self-concept to adjustment in a selected group of college women. Unpublished doctoral dissertation, Michigan State College, 1953.
- DENNIS, W. The adolescent. In L. Carmichael (Ed.), *Manual of child psychology*. New York: Wiley, 1946. Pp. 633-666.
- ENGEL, MARY. The stability of the self-concept in adolescence. Unpublished doctoral dissertation, George Peabody College, 1956.
- HALL, G. S. *Adolescence*. Vol. 1. New York: Appleton, 1904.
- JERSILD, A. T. *In search of self*. New York: Teachers' College, Columbia Univer., 1952.
- KUHLEN, R. G. Age trends in adjustment during adult years as reflected in happiness ratings. *Amer. Psychologist*, 1948, **3**, 307.
- LECKY, P. *Self-consistency: A theory of personality*. New York: Island Press, 1945.
- ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change*. Chicago: Univer. Chicago Press, 1954.
- STEPHENSON, W. Correlating persons instead of tests. *Charact. Pers.*, 1935, **6**, 17-24.
- SWEETLAND, A., & QUAY, H. A note on the *K* scale of the Minnesota Multiphasic Personality Inventory. *J. consult. Psychol.*, 1953, **17**, 314-316.
- TAYLOR, D. M. Changes in the self concept without psychotherapy. *J. consult. Psychol.*, 1955, **19**, 205-209.
- TUDDENHAM, R. D. Studies in reputation. *Psychol. Monogr.*, 1952, **66**, No. 1 (Whole No. 333).
- ULLMANN, C. A. Identification of maladjusted school children. *Publ. Hlth. Monogr.*, 1952, No. 7.

Received November 21, 1957.

A SEMANTIC ANALYSIS OF A NORMAL AND A NEUROTIC THERAPY GROUP¹

ZELLA LURIA

University of Illinois

THIS paper reports an investigation of the meanings associated with certain concepts, selected for their presumed significance to psychotherapy, by subjects in therapy as compared with subjects that have never been in therapy. The instrument used was the semantic differential, a test of meaning (Osgood, Suci, & Tannenbaum, 1957).

Moss (1953) suggested that in successful therapy meanings expressed on the semantic differential in a waking state (conscious meanings) become more congruent with meanings expressed under hypnosis (unconscious meanings). Lazowick (1955) found that on the semantic differential low anxiety Ss described their like-sexed parent as more like themselves than did high anxiety Ss. These results imply that in therapy a shift in identification may occur towards the like-sexed parent as part of the process of diminishing anxiety.

The purpose of the research reported here was to see if Ss entering therapy can be differentiated from persons that have never been in therapy by the meanings they associate with certain selected concepts and to test for areas where semantic changes followed therapy, especially as they may relate to parental identification.

METHOD

The Ss consisted of three control and three therapy groups. Control Group I (hereafter called CI) included 62 college sophomores, 9 males and 53 females. Control Group II (CII) consisted of 18 male first-year graduate students in psychology. To insure against bias due to differences in the sex ratios of the control and therapy samples, a third control group (CIII) was chosen from among the CI Ss. CIII was a sample of 16 females and 9 males matched for sex, approximate age, and,

where possible, marital status with Therapy Group TIII described below.

In order to insure that Ss in therapy were representative and not selected, even inadvertently, by their therapists to represent the problem area of the particular therapist's interest, clients were drawn from therapists of different theoretical persuasions. The first therapy group (TI) consisted of 50 Ss (32 males and 18 females) in therapy mainly in university settings. Sixteen were clients of Carl Rogers or his colleagues; 22 Ss were clients of a group of neo-Freudian therapists, primarily at the University of Illinois Counseling Bureau; 12 Ss were in therapy with O. H. Mowrer. The TI group consisted of 32 males and 18 females. A second therapy group (TII), clients of neo-Freudian therapists at the University of Illinois Counseling Bureau,² included 32 Ss, 25 males and 7 females. The third therapy group (TIII), consisting of 16 females and 9 males, was chosen from TI for matching with Group CIII as described above.

CI, CIII, TI, and TIII were tested three times at two-month intervals. The first testing for these groups consisted of a test and immediate retest. The tests were administered individually to all therapy Ss at their counseling centers. The author administered the test to the CI and CIII groups in the college classroom. The CII group also took the test in a class setting. Instructions were mimeographed on the test form. After an introductory statement by the tester to the effect that the test was to ascertain meanings, all questions were answered from the mimeographed instructions.

Retest measures for reliability estimates were obtained for CI, CIII, TI, and TIII. No reliability measures were available for CII and TII.

The test used was a form of the semantic differential, a measure of certain aspects of connotative meaning described by Osgood, Suci, and Tannenbaum (1957). The concepts were chosen from a total of 45 terms, suggested by graduate students in psychology and by some of the therapists whose clients were tested. The 15 concepts chosen fall into two general categories, significant persons and conceptual abstractions. In the former group were the concepts ME, MY MOTHER, MY FATHER, MY SPOUSE, CHILD, MY DOCTOR (or MY THERAPIST or MY COUNSELOR). In the latter group were LOVE, MY JOB, MENTAL SICKNESS, PEACE OF MIND, FRAUD, SELF-CONTROL, HATRED, CONFUSION, and SEX. These concepts met two criteria: relevance to therapy and a reasonable variability in response as indicated by a preliminary sample of Ss.

The scales were pairs of adjectives chosen by a

¹ The author is indebted to the Ford Foundation Fellowship Committee of the University of Illinois for their support of this work and to C. E. Osgood and O. H. Mowrer for their active help. Carl Rogers and the staff of the University of Chicago Counseling Center, William Gilbert and the staff of the University of Illinois Student Counseling Bureau, O. H. Mowrer, and Harriett P. Ray generously provided clients' data.

² I am indebted to William Gilbert, J. McV. Hunt, T. Ewing, and N. Endler for making data for Therapy Group II available.

combination of methods. Therapy records were examined for adjectives recurring in clients' statements. Adjectives were chosen for use as scales when they (a) tended to recur in clients' records and (b) showed high factor loadings in a factor analysis made by Osgood and Suci (1952). These authors found three factors of connotative meaning: an evaluative factor, represented here by the scales *clean-dirty*, *valuable-worthless*, and *tasty-distasteful*; a potency factor, represented here by *strong-weak*, *large-small*, and *deep-shallow*; and an activity factor, represented by *hot-cold*, *fast-slow*, and *active-passive*. A tenth scale about equally loaded on the three factors, *relaxed-tense*, was used because of its frequent occurrence in case records. Each scale contained seven scoring positions.

RESULTS

Reliability of the Semantic Differential

Under conditions of testing followed by immediate retest, CIII and TIII, the matched groups, appeared to be equally reliable when extremity of score (closeness to Scale Position 1 or 7) was taken into account. The analysis was limited to eight concepts (ME, MY MOTHER, MY FATHER, MENTAL SICKNESS, PEACE OF MIND, HATRED, CONFUSION, and SEX) because of excessively low variance on the other seven concepts, whose inclusion would have inflated the estimates of reliability.

While TIII, the experimental group, shifted upon retest an average of .79 scale units on these concepts as compared with .69 for CIII, the difference is most likely a reflection of extremity of score rather than an intrinsic difference in reliability.

Figure 1 shows the relationship of extremity of score to average shift between the first test (1a) and the immediate retest (1b). Clearly, the group that scores closer to 4, a position of neutrality or ambivalence, is the less reliable group. Analysis of data for a larger group, composed of CI and TI, yielded a significant rho of $-.81$ between average shift and extremity of score. This finding, similar to results discussed by Festinger and Katz (1953), seems simply to indicate that an intense attitude shifts less readily on retest than a less intense one.

The results shown in Fig. 1 also confirm the results of other studies with the semantic differential (Osgood et al., 1957) in indicating that the evaluative factor³ gives the most re-

³ Estimated by pooling the concepts on evaluative scales: *valuable-worthless*, *tasty-distasteful*, and *clean-dirty*.

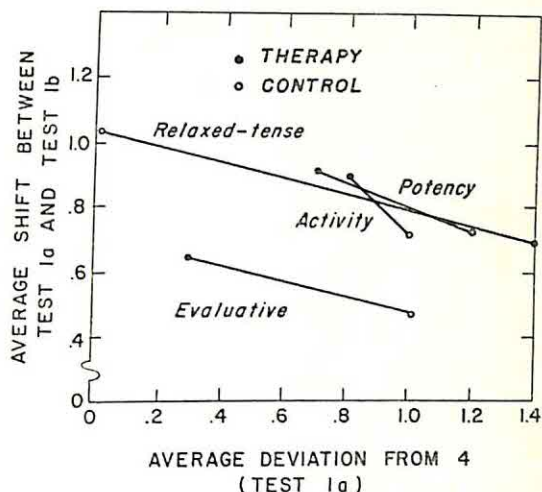


FIG. 1. RELATIONSHIP OF EXTREMITY OF SCORE (DEVIATION FROM SCORING OF 4) TO AVERAGE SHIFT IN RETESTING FOR MATCHED THERAPY AND CONTROL GROUPS: TEST 1A = FIRST TEST; TEST 1B = RETEST

liable results. Average shifts go from a low of .54 scale units on the evaluative factor, through the .80's on the activity factor,⁴ to .90 on the potency factor.⁵

Two conclusions seem justified: (a) The evaluative factor is consistently the most reliable, and (b) extremity of scores has a positive effect on reliability.

Comparison of Therapy and Control Groups

When the average pretherapy scores on the three factors and on the scale *relaxed-tense* are analyzed for the eight concepts described above, no significant differences appear between CIII and TIII in their judgments of MENTAL SICKNESS, PEACE OF MIND, HATRED, CONFUSION, or SEX. The only concepts showing significant differences between the matched groups are ME, MY MOTHER, and MY FATHER. For these three concepts, on all three factors and on *relaxed-tense*, the TIII scores are on the average closer to the unfavorable (*bad*, *weak*, *passive*, and *tense*) end of the scales than the CIII scores.

Thus, ME for TIII was rated on the average for the evaluative factor at 3.4, for the potency factor at 3.7, for the activity factor at

⁴ Estimated by pooling the concepts on activity scales: *active-passive*, *hot-cold*, *fast-slow*.

⁵ Estimated by pooling the concepts on potency scales: *strong-weak*, *deep-shallow*, and *large-small*.

TABLE 1

CHI SQUARE ANALYSES ON THE CONCEPTS ME, MY MOTHER AND MY FATHER FOR MATCHED CONTROL AND THERAPY GROUPS

Factor	Group	ME		MY MOTHER		MY FATHER	
Evaluative	T C	<i>1 and 2</i>	<i>3 and over</i>	<i>1</i>	<i>2 and over</i>	<i>1</i>	<i>2 and over</i>
		5	20	6	19	6	19
		18	7	19	6	16	9
		$\chi^2 = 13.6^{**}$		$\chi^2 = 13.5^{**}$		$\chi^2 = 8.1^*$	
Potency	T C	<i>1 and 2</i>	<i>3 and over</i>	<i>1 and 2</i>	<i>3 and over</i>	<i>1 and 2</i>	<i>3 and over</i>
		4	21	2	23	8	17
		14	11	15	10	20	5
		$\chi^2 = 8.7^*$		$\chi^2 = 15.0^{**}$		$\chi^2 = 11.6^{**}$	
Activity	T C	<i>1 and 2</i>	<i>3 and over</i>	<i>1 and 2</i>	<i>3 and over</i>	<i>1 and 2</i>	<i>3 and over</i>
		9	16	11	14	3	22
		18	7	16	9	13	12
		$\chi^2 = 6.5^*$		$\chi^2 = 2.0$		$\chi^2 = 9.3^{**}$	
Relaxed-tense	T C	<i>1 to 4</i>	<i>5 to 7</i>	<i>1 to 4</i>	<i>5 to 7</i>	<i>1 to 4</i>	<i>5 to 7</i>
		2	23	2	23	8	17
		18	7	19	6	14	11
		$\chi^2 = 21.3^{**}$		$\chi^2 = 23.8^{**}$		$\chi^2 = 2.9$	

* Significant at .025 level.

** Significant at .005 level.

Note.—Scores are given in italics, frequencies in Roman figures.

3.4, and for *relaxed-tense* at 6.1. CIII's ratings were 2.3, 2.8, 2.5, and 3.3, respectively. Differences in the same direction and of the same order appear for the concepts MY MOTHER and MY FATHER. These differences persist when group CII is added to group CI, and the combined group is compared with Group TI.

Table 1 shows the distribution of scores for CIII and TIII on the three factors and on the *relaxed-tense* scale for ME, MY MOTHER, and MY FATHER. Chi square values for the differences between the two matched groups are shown in Table 1. It is clear that the control Ss scored the three concepts in a significantly more favorable way than did the therapy Ss. Ten of the 12 differences, including all those on the evaluative and potency factors, are significant at or beyond the .025 level. The two nonsignificant differences are in the same direction as the significant ones.

Frequency of Extreme and Central Scores

An analysis was made of the use of the extreme scores by CIII and TIII at the first testing. The purpose of the analysis was to check Johnson's hypothesis (1946) that neurotics tend to think in dichotomies—for ex-

ample, in terms of either failure or success, good or bad—with no intermediate judgments. The results showed that Scale Positions 1 and 7 were used on the average in 41% of the scores by CIII and in 38% of the scores by TIII. This finding does not support Johnson's contention.

On the other hand, Mowrer has suggested (1953) that the neurotic process is set in motion by guilt due to evasiveness. With a score of 4 taken as an index of evasion because it seems to reflect indiscriminateness, ambivalence, or lack of opinion, the two groups were compared. The therapy group used Score 4 in 12% of their scores; controls used it 17% of the time. This finding fails to support Mowrer's hypothesis, but the incidence of 4 may be an inadequate measure of evasiveness.

Parent Identification Analysis

The problem of identification with parent figures plays a part in almost all theories of therapy. Mowrer has hypothesized (1953) that the neurotic suffers from faulty identification and often identifies with the parent of the opposite sex. In order to test if therapy Ss differ before therapy from nontherapy Ss in choice of favored parent, a chi square analysis was made comparing CI and TI for their scores on the evaluative factor for the concepts MY MOTHER and MY FATHER. The larger groups, rather than the matched groups, were used in order to reach the necessary cell frequencies for the statistical analysis.

The control and therapy groups both favored the mother, in agreement, apparently, with the over-all cultural pattern in American society. In the therapy group, 45% rated their fathers as less valued than their mothers; in the control group, 38% held the same view. The therapy group rated the father more favorably than the mother in 31% of the cases, compared with 14% in the control group. A striking characteristic of the control groups, both CI and CII, was that almost half of them (48%), regardless of sex, scored the two parents identically on the evaluative factor, whereas only 24% of the therapy Ss rated their parents as equally "good." The chi square for the difference between the distributions of the two groups shows significance beyond the .01 level.

An attempt was made by a pattern analysis of the evaluative scores for ME, MY FATHER, and MY MOTHER to see if the therapy Ss showed

any configuration that would distinguish them from normal controls. No clear neurotic paradigm was found for either of two patterns: (1) "I'm fine. My parents aren't" (ME judged "better" than both parents), or (2) "They're fine. I'm not" (ME judged "worse" than both parents). There was also no evidence to indicate that Ss who rate themselves as superior or inferior to their parents discriminate more widely between their parents.

Process of Change Following Therapy

The process of change following therapy was analyzed by a comparison of 25 Ss selected from TII with all Ss in Group TIII. This procedure was required because of the small size of the posttherapy sample from TIII. The 25 Ss were selected from among the original 32 in Group TII by a matching of TII and TIII Ss on pretherapy factor scores for a given concept in order to control differences attributable to differences in initial scores.

Difference scores were obtained for TII by subtracting pretherapy from posttherapy factor scores. These difference scores were compared with the difference scores for TIII obtained by subtracting first test factor scores from the immediate retest scores.

This analysis yielded *t* values for the differences between matched samples that were significant only for the concept ME. Thus, the group that concluded therapy, TII, showed a significant improvement in its judgment of the self concept after therapy. ME changed for TII on the evaluative factor an average of 2.1 scale units, as compared with only .9 for TIII; on the potency factor, TII shifted an average of 2.1 scale units as compared with a shift by TIII in a devaluing direction of $-.2$; on *relaxed-tense*, TII shifted an average of 1.2 scale units as compared with TIII's shift of $.2$. While TII's activity factor score was not significantly different from TIII's, the analyses of the other two factors and of *relaxed-tense* were significant at the .02 level.

DISCUSSION

The findings indicate that control Ss can be differentiated from therapy Ss prior to therapy by the meanings they attribute to the concepts of the self and of parents on the semantic differential. Therapy Ss devalue these concepts; control Ss apparently do not.

That therapy affects the self-estimate and not the judgment of the parents may seem surprising. When one recalls, however, that the therapy Ss were predominantly away from home and family, being treated in university settings, one may have a partial explanation. If therapy is relearning or reality-testing with new meanings learned in therapy, the relevant persons may have to be present. Of these, only the self was available to our sample; the parents were not. Furthermore, clients may tolerate more easily a devaluing picture of their parents (or perhaps an accurate perception?) than a devaluing perception of themselves. It was conjectured by Osgood and Luria (1954) that the dissociation in a multiple personality case came in part from the patient's striving for an acceptable estimate of herself.

The present results are in agreement with those of Rogers & Dymond and others (1954) on the central role of the self concept and suggest that further refinement of the self concept may provide potent tools for the study of personality development.

The result on comparison of the parents suggests that differential evaluations of the two parents do not occur in most normal Ss. Perhaps normality requires positive acceptance of both parent roles. Thus, acceptance and valuation of both parents may be more important for the development of successful identification than discrimination between the roles of the two parents.

The influence of the response tendencies of the control and therapy Ss on the results should not be overlooked. The meaning of the self and the parents for the therapy Ss, before therapy, is most frequently in the "greys." Control Ss tended, instead, to use "black and white." This may influence measurements of semantic distance among the three concepts. But the fact remains that the controls were not self- and parent-devaluing, whereas the therapy Ss were. Devaluation of self and parents may well be diagnostic of people in trouble; improvement in therapy may require improvement of the self concept, not necessarily of parent concepts.

Results of pretherapy testing did not reveal any significant differences among the three groups selected from different therapists. This may either be due to small sample size or to a true lack of selection on the part of the therapists.

pist. Semantic studies of the effects of therapy in general and specifically by therapists of differing persuasions would be very desirable. The method used here could provide valuable information.

Many questions are raised for further study. What role does the change in the self concept play? How does the change in the self concept correlate with the success of the therapy? Does the successful client shift semantically toward greater agreement with his therapist's semantic framework? Therapy may well prove to be a process of reinforcement for semantic agreement by the client with the therapist's meaning system as suggested by Greenspoon (1955) and Verplanck (1955). A central problem may be that of how behavioral changes become associated with semantic changes.

SUMMARY

The reliability of the semantic differential has been reported for a group of Ss entering therapy and for a control group. The evaluative factor was found to be highly reliable; potency and activity were less reliable factors for both groups.

The therapy group could be differentiated from the nontherapy group on the basis of the meanings of certain key personal concepts, self and parent figures. The therapy Ss rated these concepts less favorably than nontherapy Ss. Therapy Ss tended to be somewhat less extreme than control Ss in their semantic judgments.

About half the control Ss view the parent figures as being alike on the evaluative factor as compared with only one-fourth of the

therapy Ss. Where any distinction is made between the parents, the mother tends to be the more favorably rated one for both therapy and control Ss.

The therapy process appeared to improve the judgment of the self concept significantly. Judgments of the parent concepts did not seem to change.

REFERENCES

- FESTINGER, L., & KATZ, D. *Research methods in the behavioral sciences*. New York: Dryden, 1953.
- GREENSPOON, J. The reinforcing effect of two spoken sounds on the frequency of two responses. *Amer. J. Psychol.*, 1955, **68**, 409-416.
- JOHNSON, W. *People in quandaries*. New York: Harper, 1946.
- LAZOWICK, L. On the nature of identification. *J. abnorm. soc. Psychol.*, 1955, **51**, 175-183.
- MOSS, C. S. An experimental investigation of symbolic dream processes. Unpublished doctoral dissertation, Univer. of Illinois, 1953.
- MOWRER, O. H. (Ed.) *Psychotherapy: Theory and research*. New York: Ronald, 1953.
- OSGOOD, C. E., & LURIA, ZELLA. A blind analysis of a case of multiple personality using the semantic differential. *J. abnorm. soc. Psychol.*, 1954, **49**, 579-591.
- OSGOOD, C. E., & SUCI, G. J. A measure of relation determined by both mean difference and profile information. *Psychol. Bull.*, 1952, **49**, 251-262.
- OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana, Ill.: Univer. Illinois Press, 1957.
- ROGERS, C. R., & DYMOND, ROSALIND F. (Eds.) *Psychotherapy and personality change: Coordinated studies in the client-centered approach*. Chicago: Univer. Chicago Press, 1954.
- VERPLANCK, W. S. The control of the content of conversation: Reinforcement of statements of opinion. *J. abnorm. soc. Psychol.*, 1955, **51**, 668-676.

Received December 2, 1957.

MANIFEST ANXIETY AND REVERSED ALPHABET PRINTING

JANET A. TAYLOR

Northwestern University

AND ALLAN RECHTSCHAFFEN¹

University of Chicago, Medical School

A STUDY was recently conducted by one of the present writers (Rechtschaffen, 1958), part of which involved testing a group of subjects (Ss) on a reversed alphabet printing task under both massed and distributed conditions. The present report concerns a reanalysis of these performance data to determine their relationship to scores on the Manifest Anxiety Scale (MAS). Several alternative predictions could be made with respect to the nature of this relationship. According to the hypotheses of Spence (1956), Taylor (1953, 1956), and their associates, who initiated the series of studies relating the MAS to performance on certain learning tasks, scores on the MAS are considered to be related in part to Ss' level of emotionality or drive. Applying certain notions from Hullian theory, these writers have postulated an interaction between drive level and learning task variables, higher drive levels facilitating performance in simple tasks in which a single response tendency is being acquired and inhibiting it on tasks in which a number of intratask responses are evoked and the correct response is relatively weak. Considerable response interference appears to be involved in reversed alphabet printing, the highly practiced patterns of the letters in their ordinary form often interfering with the production of the reversed letters. A negative relationship would therefore be expected between task performance and scores on the MAS, high scorers (high drive Ss) being inferior to low scorers.

In contrast to the drive interpretation of anxiety phenomena are the hypotheses proposed by Eysenck (1955) in connection with his work on the introversion-extroversion dimension. Extending certain notions put forth by Pavlov, Eysenck has proposed that reactive inhibition, as it is defined by Hull (1943), is

generated more easily and dissipates less quickly in extroverted individuals than in introverts. He has further stated that the MAS is, in part, related to the introversion-extroversion dimension, high scorers on the MAS tending to be introverts. Thus, Eysenck believes, the demonstrated superiority of "high anxiety" groups (as defined by the MAS) in such tasks as classical conditioning may be attributed to the development of lesser amounts of reactive inhibition rather than to a higher drive level. In the present task, Eysenck's hypothesis would predict, contrary to drive theory, an over-all *positive* relationship between MAS scores and task performance (particularly during massed practice), the low anxious Ss becoming increasingly inferior to the high anxious with successive trials due to the greater accumulation of reactive inhibition.

Additional analyses of the data were also made. A factor analysis performed by O'Connor, Lorr, and Stafford (1956) has identified five different factors in the MAS. Each of these factors could, of course, be related to performance in the same manner. If one or more were not, however, a number of practical and theoretical problems affecting MAS research would arise. Five MAS subscores, each consisting of the items assigned to them by O'Connor et al., were therefore obtained for each S and related to task performance in the same manner as scores on the total test.

METHOD

Subjects. The 96 Ss, 48 men and 48 women, drawn from an introductory psychology class, were originally selected in such a manner that half fell above and half below the class mean on the R (Rhathymia) Scale of Guilford's Inventory of Factors STDCR (Guilford & Guilford, 1939). Since scores on the R scale were not significantly correlated with scores on the MAS (which had been given to the class in another connection), the Ss can be considered to have been unselected with respect to the latter.

MAS subscores. In attempting to identify the parameters underlying each of the five extracted fac-

¹ The major portion of this research was done while the writer was at the Chicago Branch of the VA Neuropsychiatric Research Laboratory at the West Side VA Hospital, Chicago, Illinois.

tors, O'Connor, Lorr, and Stafford considered only factor item correlations of .35 or higher; this practice was followed in the present study in determining subscores. Of the 38 (out of 50) items that met this level, eight reached it for two of the factors and were thus included in two of the subscores. The exact items involved and their factor loadings may be found in the original report of O'Connor et al. (1956, p. 161). The description of the factors, as offered by these writers, is as follows: A—(12 items) chronic anxiety associated with self-consciousness, sensitivity, and lack of personal security; B—(8 items) increased physiological reactivity to emotional stimuli (e.g., blushing, sweating, easy embarrassment, etc.); C—(5 items) intense inner strain associated with sleep difficulty; D—(8 items) a sense of personal inadequacy; E—(14 items) motor tension (restlessness, tiring quickly, etc.).

Experimental procedure. Each *S*, tested individually, was instructed to print the alphabet upside down and backwards in a test booklet as rapidly as possible. The *S* was given five 30-sec. trials separated by 30-sec. rest intervals, a 1-min. rest interval, and finally 5 min. of continuous practice. During the latter, *S* turned to the next page every 30 sec. at a signal from *E*, thus permitting analysis of performance in successive 30-sec. units.

RESULTS

The clearest demonstration of the relationship between MAS scores and task performance, particularly with respect to the development of reactive inhibition, is afforded by comparing the performance of *Ss* selected from the extremes of the MAS distribution to form a high and low anxiety group. The 32 *Ss* (lower third) with the lowest MAS scores (10 or below) and the 32 (upper third) with the highest scores (17 or above) were chosen for this purpose. The mean number of letters

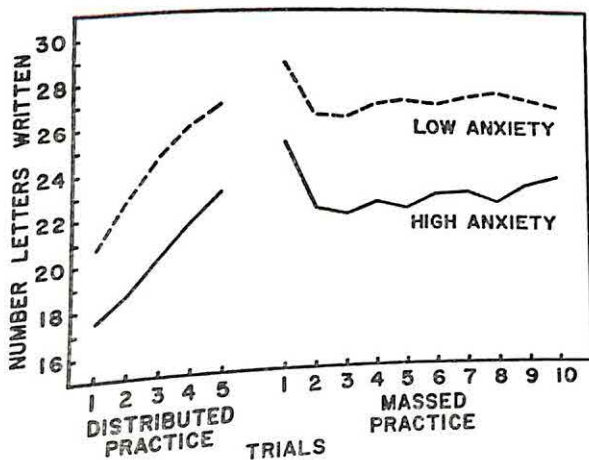


FIG. 1. MEAN NUMBER OF LETTERS PER TRIAL DURING MASSED AND DISTRIBUTED PRACTICE FOR THE TWO EXTREME ANXIETY GROUPS

TABLE 1
ANALYSIS OF VARIANCE FOR REPEATED MEASURES OF THE PERFORMANCE OF THE TWO EXTREME GROUPS DURING DISTRIBUTED AND MASSED PRACTICE

Source	df	Distributed		Massed	
		Mean Square	F	Mean Square	F
Between groups	1	1304.11	10.62*	2877.26	14.07*
Between <i>Ss</i> same group	62	122.83		204.50	
Between trials	4	373.90	84.59*	39.97	10.57*
Trials \times groups	4	5.23	1.18	4.79	1.27
Pooled <i>Ss</i> \times trials	248	4.42		3.78	

* Significant at .01 level of confidence.

TABLE 2
OBTAINED AND ESTIMATED CORRELATION COEFFICIENTS (WITH LENGTH OF TEST HELD CONSTANT) BETWEEN PERFORMANCE AND VARIOUS MAS MEASURES

MAS Measure	No. Items	$r_{obt.}$	$r_{est.}$
Factor A	12	-.218*	-.226*
B	8	.026	.030
C	5	-.156	-.289**
D	8	-.127	-.146
E	14	-.218*	-.218*
Rejected items	12	-.204*	-.212*
Total scale	50	-.272**	

* Significant at .05 level.

** Significant at .01 level.

printed by the two groups in each 30 sec. trial during both the distributed and massed series are shown in Fig. 1. As may be seen, the low anxiety group was superior in performance throughout. Further, although the massed practice session seemed to result in the development of considerable reactive inhibition, as indicated by the downward slopes of the curves, the magnitude of the difference between the two groups appeared to remain constant.

The data for the distributed and massed trials were each subjected to an analysis of variance for repeated measures (Edwards, 1950), the results of which are summarized in Table 1. In each case the between-groups *F* was significant (at the .01 level or beyond), whereas the interaction between groups and trials was not. Since there was no indication of a differential effect of the two types of practice on the extreme groups, a single performance measure, the total number of letters printed in all trials, was obtained for each *S*. The Pearson *r* be-

tween this measure and the MAS, using all 96 Ss, was $-.272$, a value which is significant at the .01 level of confidence.

Correlations were also computed between task performance and each of the subscales corresponding to the five O'Connor factors. As a matter of incidental interest, a correlation was also obtained between performance and scores on the 12 items that were not included in any of the factors. As indicated in Table 2, all r 's except that for Factor B were in the expected negative direction. In spite of the small number of items on each subscale, three of them (for Factors A, E, and the 12 rejected items) were significantly different from zero at the .05 level. The different correlation values were due in part, of course, to the different number of items on each subscale. Using an equation suggested by Gulliksen (1950, Equation 12, p. 94), an attempt was made to compare the goodness of the subscales more directly by estimating the r 's that would have been obtained if each had had the same number of items as the longest subscale (14 items). Table 2 shows that little change in the value of the r 's resulted from this procedure except in the case of Factor C, the shortest of the subscales, which shifted from $-.156$ to $-.289$.

DISCUSSION

The results of the present study confirm the expectation drawn from the drive interpretation of anxiety that MAS scores would be inversely related to task performance. Eysenck's hypotheses, however, were not confirmed; not only was there no evidence that differential amounts of reactive inhibition developed in the two groups during massed practice but also the group differences that did obtain were in the opposite direction to those predicted. The failure to support Eysenck's hypotheses may reflect the lack of correlation between the MAS and his introversion-extroversion dimension, between the latter and rate of development of reactive inhibition, or both. Scores of the Ss on the Guilford Rhythymia Scale (1939), often used by Eysenck and his colleagues to measure the extroversion dimension, did not, as mentioned earlier, correlate significantly with the MAS ($r = -.188$) nor were they significantly related to various measures of task performance, including a measure of the amount of reactive inhibition

generated by massed practice (Rechtschaffen, 1958).

With respect to the separate factors on the MAS, O'Connor, Lorr, and Stafford (1956) have questioned the drive interpretation of MAS research on the basis of their findings: "More or less implicit, at least for users of the A scale, is the assumption that the scale measures a simple unitary drive. . . . If the Scale is to be used as a systematic measure of drive or motivational level in experimental work, it is important to determine whether one or more drives are manifested in the scores" (p. 160). In a recent review, Eriksen (1957) also cast doubt on the use of the MAS in drive research because of its multidimensionality. Such statements seem to be based on a misunderstanding of the drive construct as it is employed in the Hullian system: there is, in essence, but a single drive state, its value being determined by the aggregate strength of all primary and secondary needs, relevant or irrelevant, that are operative at the moment (Hull, 1943; Spence, 1956). Thus, even if it is assumed that each of the factors represents a different source of drive, the multidimensionality of the MAS has no necessary significance as far as drive theory is concerned, as long as each dimension may be shown to have drive properties.² The results of the analyses of the subscales, including the one containing the 12 items not assigned to any factor, suggest that all but one have the postulated drive properties, i.e., were related to the experimental criterion in the expected manner. Since the eight items constituting the deficient factor were essentially unrelated to the experimental measure, a more efficient test may well result from the omission of these from the total scale.

² While the multidimensionality of the MAS has no necessary theoretical significance for drive research, a practical question concerning subject selection might arise if one were to assume that a combination of different drive sources was being reflected by total scores on the test. For example, might a person receiving a very high score on one of the factors and 0 on the rest better be classified as a "high drive" S rather than as a relatively "low drive" S, as would be the case with current procedures? However, an examination of the O'Connor, Lorr, and Stafford factor matrix (1956, p. 161) shows that it is largely positive, indicating that score discrepancies among the factors, such as illustrated above, would be very unlikely to occur.

SUMMARY

A group of 96 Ss was given a reversed printing task, presumed to involve considerable intratask interference. Task performance was found to be negatively correlated with Ss' scores on the Manifest Anxiety Scale. Scores on a number of subscales, corresponding to factors extracted from the total scale in a previous study, were also computed for each S and these correlated with performance. All but one of these r 's were in the expected negative direction. The theoretical implications of these findings were discussed.

REFERENCES

- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rhinehart, 1950.
- ERIKSEN, C. W. Personality. In P. P. Farnsworth (Ed.), *Annual review of psychology*. Stanford, Calif.: Stanford Univer. Press, 1957. Pp. 185-210.
- EYSENCK, H. J. Cortical inhibition, figural aftereffect, and personality. *J. abnorm. soc. Psychol.*, 1955, **51**, 94-106.
- GUILFORD, J. P., & GUILFORD, R. B. Personality factors, D, R, T, and A. *J. abnorm. soc. Psychol.*, 1939, **34**, 21-36.
- GULLIKSEN, H. *Theory of mental tests*. New York: Wiley, 1950.
- HULL, C. L. *Principles of behavior*. New York: Appleton-Century, 1943.
- O'CONNOR, J. P., LORR, M., & STAFFORD, J. W. Some patterns of manifest anxiety. *J. clin. Psychol.*, 1956, **12**, 160-163.
- RECHTSCHAFFEN, A. Neural satiation, reactive inhibition, and introversion-extraversion. *J. abnorm. soc. Psychol.*, 1958, **57**, 283-291.
- SPENCE, K. W. *Behavior theory and conditioning*. New Haven: Yale Univer. Press, 1956.
- TAYLOR, JANET A. A personality scale of manifest anxiety. *J. abnorm. soc. Psychol.*, 1953, **48**, 285-290.
- TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303-320.

Received December 2, 1957.

INTRINSIC MOTIVATION AND PSYCHOLOGICAL STRESS¹

WILLIAM VOGEL, SUSAN RAYMOND

Clark University

AND RICHARD S. LAZARUS

University of California, Berkeley

IN THE theoretical approach to psychological stress recently proposed by Lazarus and Baker (1956, 1957), two problems were emphasized. On the one hand, an individual's pattern of motivation was regarded as determining the potency of any situation in producing stress. On the other hand, once a stress reaction is aroused, the person's behavior depends upon his method of coping with the disturbance. Factors both of motivation and of cognitive control therefore need to be included in the systematic study of psychological stress.

Some experimental efforts to examine this point of view have recently been reported (Lazarus, Baker, Broverman, & Mayer, 1957; Vogel, Baker, & Lazarus, 1958), mainly involving the manipulation of motive strength both by its experimental arousal and by selection of *S*s with differing degrees of some relatively enduring motivational trait (intrinsic motivation). The present paper is mostly restricted to the experimental analysis of the role of intrinsic motivation.² It deals with three main issues.

First and most important is the role of type of motivation in determining whether a stress state occurs. An excellent recent review of research in the area of ego involvement has been presented by Iverson and Reuder (1956). We propose that any typical stressor situation will produce a state of stress only if it threatens important ego-motivations of the *S*. Thus, if the experiment casts doubt on the intellectual adequacy of an *S* who has little motivation to achieve intellectually, there will be little or no stress. In this study we selected *S*s who were primarily either achievement motivated or affiliation motivated, exposing them to an

achievement stressor and an affiliation stressor condition. Our expectation was that a state of stress would occur only if the stressor condition communicates with the predominant motive pattern.

The second issue concerns the relation of the stress state to task performance. The problem is a complicated one, and the evidence from the literature suggests that there is no simple relationship between stress and performance (Lazarus, Deese, & Osler, 1952; McClelland, 1951); sometimes there appears to be facilitation of performance and at other times there is impairment. It has been shown that facilitation or impairment is partly a function of the *S*'s cognitive structure (Broverman & Lazarus, 1958). Others (Sarason & Mandler, 1952) have suggested that one should expect no change, facilitation, or impairment in performance, respectively, depending upon the compatibility of the stressor-produced responses with the required task operations. Our present approach to the problem includes the use of two different types of tasks, a perceptual-motor task (the McKinney Reporting Test: McKinney, Strother, Hines, & Allee, 1951), and a more conceptual task (scrambled words). We hypothesized that a stress state would increase perceptual motor output (a frequent finding in other research [Lazarus et al., 1952]) but would impair conceptual operations. In the case of the former, increased affect should be compatible with greater motor output, while in conceptual operations, the affect should result in interference.

Finally, we chose one infrequently used additional variable, the *S*'s past history of success or failure in his predominant motivational sphere, that is, achievement or affiliation, because such success or failure might determine the nature of the *S*'s apprehension of the present stressor situation, or might reflect personality characteristics associated with the ability to master stressful experiences. Previously successful *S*s, on both above bases,

¹ This investigation was supported by a research grant (M-734) from the National Institute of Mental Health of the National Institutes of Health, Public Health Service.

² This article is the first of two which actually represent a single experimental effort to study simultaneously the role of intrinsic motivation and cognitive control factors in psychological stress.

should be less readily disturbed by the stressor conditions.

METHOD

Selection of Subjects

The assessment of achievement motivation and of affiliation motivation was accomplished by means of a variety of techniques which were devised to screen a large group of Ss, selecting for the experiment only those with clear manifestations of the desired motive traits. The selection plan involved an attempt to obtain ideal cases who were either highly motivated toward achievement and disinterested in affiliation, or highly motivated toward affiliation and disinterested in achievement. These Ss were further subdivided into successful and unsuccessful groups so that four groups of 10 Ss each were ultimately created: affiliation oriented³ successful, affiliation oriented unsuccessful, achievement oriented successful, and achievement oriented unsuccessful. Let us consider the six techniques devised to select Ss with these characteristics.

Academic achievement in relation to aptitude. This index was based on the S's academic class standing in relation to his performance on the American Council on Education college aptitude test. Both distributions were expressed in standard score form, and the difference between the paired scores identified some Ss whose aptitude was greater than achievement (under-achievers) and some who were relatively high in academic achievement compared with aptitude (over-achievers).

Amount of study. Each S was required to indicate on a chart, on which half-hour intervals each day for a whole week were arrayed, what he had done during these intervals the previous week. Each half-hour space had to be filled in. To serve as a guide, a number of activities were illustrated on the margin of the chart, including organized school athletics, dating, school club activity, studying, relaxing alone, bull sessions, cards, etc. From this it was possible to obtain information about the amount of time the S spent studying or in some closely related activity, and the amount of social activity he had engaged in.

Values questionnaire. Each S was given a 15-item test, each question to be answered true or false. The test was constructed so that achievement values were pitted against affiliation values. For example, Item 4 stated, "Competition is bad, it kills friendship";

Item 8 was, "I would rather be voted 'most likely to succeed' than 'most popular guy on campus'"; and Item 11 read, "People who study a lot seem pretty one-sided, they don't have too many friends." The answer "true" to Items 4 and 11 suggests stronger affiliation than achievement values, while a "true" for Item 8 suggests the reverse.

Social attainment scale. A shortened adaptation of the Worcester Social Attainment Scale (Phillips & Cowitz, 1953) was prepared on which the S was required to indicate memberships in social organizations, teams or school organizations, officerships held, jobs held, community activity, and so on. It was possible by means of this questionnaire to estimate the degree of accomplishment of Ss in affiliative activities compared with activities that were aimed at academic or occupational achievement.

Sociometric scale. Since all of the selection techniques were administered in group, it was possible to ask each boy, in turn, to stand while the others rated him on a five-point scale concerning how well they knew him and how well they liked him. Thus, we could determine (a) how well an S was known by his peers, (b) how well he was liked by his peers, (c) how well he himself knew others in the class, and (d) to what extent he liked them.

Teachers' ratings. Each S had been asked to give the names of three teachers whom he felt knew him best. Each of the teachers was given a scale upon which he rated the S's aptitude, his academic effort, his desire for warm affiliative relationships with others, and his success at achieving warm affiliative relationships. The categories on the scale were arranged so that they represented different degrees of affiliative and achievement striving and the successfulness of the S in satisfying these needs.

During the initial selection procedures, 185 students at the Worcester Academy, a private college preparatory school in Worcester, Massachusetts, were used as Ss.⁴ The boys had been asked to cooperate by their school authorities. They ranged in age from 16 to 19 years with an occasional student over 20. They were told that they were to be tested by two Clark University psychologists but were given no explanation by the school regarding the purpose of the tests. They were initially tested in three groups of just over 60 boys each, all on the same morning. The group sessions lasted approximately 90 minutes. The Ss were told the purpose of the testing was to select a small number of students who were to be used in further experimentation.

⁴ We wish to express our great appreciation to William S. Piper, Headmaster of the Worcester Academy, and R. L. W. Smyth, Assistant in Admissions, for their valuable assistance in conducting this study at their school.

³ Henceforth we shall use the term "oriented" to refer to the groups designated in terms of the behavioral criterion of motivation, and "motivation" when we are clearly referring to the inference of a motivated state or disposition.

The six techniques for assessing achievement as opposed to affiliation motivation were utilized in such a way as to produce 40 experimental *Ss* who represented as pure cases as could be obtained. Twenty achievement oriented and 20 affiliation oriented *Ss* were selected, with half of each of these groups being the most successful or the most unsuccessful individuals within their particular motive pattern. The distribution for each trait was converted into standard scores, and *Ss* for each subgroup were taken from the upper parts of the distribution on each relevant trait.

For example, to be in the achievement oriented, successful group, the *S* had to be in at least the top 25% of the group on the criterion of the study hours, as well as in the upper 25% on the index of overachievement. In addition, the teachers had to agree that these *Ss* worked hard and, in their judgment, achieved a performance level in school commensurate with or beyond their ability. Disagreements in the teachers' ratings, or between criteria, required discarding the case. *Ss* in this group were also at the bottom half of the distribution with respect to evidence of affiliative orientation as displayed in the sociometric scale and in the teachers' ratings of social interest as well as in evidence of social activity on the social attainment scale. Finally, such an *S* was also required to be in the upper 25% in terms of achievement orientation on the self-report values questionnaire. The ten *Ss* who best fitted these criteria were selected for the achievement oriented, successful group, and the other cases were discarded. Each additional group was selected in the same general fashion.

The intercorrelations between the various indices may be examined for the light they throw on the validity of the preselection procedures. The inference that an *S* has strong affiliative motivation, for example, does not depend upon there being a high relationship between the criteria, since each may contribute some independent source of variance. For one *S*, the main avenue for attempted gratification of the motive may be in social club activities, while for another it may rest with informal conversation in the dormitories. Nonetheless, it would be supportive to our assumptions if there were a sensible pattern of relationships between the indices. Table 1 shows the pattern

TABLE 1
INTERCORRELATIONS BETWEEN CRITERIA FOR THE
SELECTION OF AFFILIATION VERSUS ACHIEVEMENT
MOTIVATION

	B	C	D	E	F	G	H
A	+.629*	+.391	+.171	+.023	-.055	+.184	-.021
	+.559	+.272	+.375	-.466	-.328	-.035	-.256
B		+.265	+.269	+.051	-.016	+.059	-.058
		+.014	+.137	-.234	-.376	+.099	-.182
C			+.364	-.041	+.087	+.231	-.091
			+.617	-.281	-.212	+.354	-.376
D				-.030	-.200	+.043	+.089
				+.268	-.227	+.199	-.444
E					+.073	+.056	+.161
					+.490	+.094	+.426
F						+.091	+.165
						-.143	+.481
G							-.097
H							-.281

Note.—Key:

- A How well *S* said he knew others, on sociometric rating test
- B How well *S* said he liked others, on sociometric rating test
- C How well others knew a subject, on sociometric rating test
- D How well others liked a subject, on sociometric rating test
- E Achievement-affiliation values test (see write-up on test construction).
- F Study hours
- G Ratings on social attainment scale
- H Overachievement index

* When $N = 40$, an r of .312 is significant at the .05 level, and an r of .403 is significant at the .01 level. When $N = 161$, an r of .169 is significant at the .05 level, and an r of .224 is significant at the .01 level. The italicized r 's are based upon an N of 161. The non-italicized r 's are based upon the 40 selected subjects.

of intercorrelations, excluding the teachers' judgments which could not be used for correlation purposes because of their particular form (a pattern analysis was needed for this index rather than a scale). The r 's are presented separately for the 40 selected cases, and for the total sample of usable records with an N of 161.

Space limitations prevent an elaborate discussion of these interesting intercorrelations. A brief examination of the table, however, reveals that there are significant correlations between the criteria, and that those which were employed to differentiate affiliative and achievement motivation tend to be negative, while those which were used within a motive

pattern are positively related. The pattern of intercorrelations gives some empirical support to the assumptions underlying our use of the various criteria to select ideal types in the affiliative versus achievement directions.⁵

Physiological Measures of the Stress State

The main criterion of the arousal of stress was based upon autonomic changes usually associated with emotions. Ss were told at the beginning of the experiment that the relationship between bodily activity and work was being investigated, and that during the experiment their pulse, blood pressure, and galvanic skin response were to be measured. Approximately five minutes was spent adapting the S to the measuring devices and chatting amiably. Measures of the physiological reactivity were recorded periodically in all three of the eight-minute experimental conditions, GSR readings taken at the first and third minutes of any particular condition, pulse at two minutes, and systolic and diastolic blood pressure at four minutes. This same procedure for all three measures was repeated twice in each of the three experimental conditions. The apparatus and method used in obtaining the physiological measures was identical to a previous published study exploring a similar problem (Vogel et al., 1958).

With respect to the statistical treatment of the physiological data, regression equations of the two experimental conditions, respectively, on the practice condition were employed

⁵ We have not included the McClelland type (McClelland, Atkinson, Clark, & Lowell, 1953) of fantasy *n* achievement as a preselection criterion because the fantasy measure is subject to the influence of many internal as well as external conditions, leaving the exact status of the measure in doubt (Lazarus, Baker, Broverman, & Mayer, 1957). We did present some of the TAT pictures to the 40 preselected Ss, once under "neutral conditions" before the experiment proper, and then again under "aroused conditions" following the stressor condition. None of the relationships between these means and any of the dependent variables in the study are significant. Moreover, few significant relationships with other preselection criteria appeared. The two significant relationships were with the social attainment scale and aroused *n* achievement ($r = +.33$), and the overachievement index and aroused *n* achievement ($r = -.38$). Each of these relationships are, in a sense, in the opposite direction from expectation. It seems to us that insufficient attention has been paid to the conditions under which the *n* achievement score measures achievement motivation

according to formulas provided by Lacey (1956), thus removing the effect of the base level (obtained under the practice condition). It is our belief that such a procedure is the most suitable one for the present experimental design, and is probably superior to the one reported in our previous research (Vogel et al., 1958).

Aside from the problem of base level, a second issue in simultaneously dealing with several measures or channels of physiological response concerns the means of combining measures to provide a single index of physiological reactivity. One technique, recommended by Lacey, involves taking the most reactive physiological channel for each S across all conditions and employing that channel as his index of physiological reactivity. A second alternative is to average the standard scores for each measure, systolic and diastolic blood pressure, pulse, and GSR, for all three conditions. The third alternative is a variant of the first in which the most reactive measure in each condition is retained for analysis, thus occasionally employing a different physiological channel for the same S under different conditions. The first and third techniques conform best to Lacey's (1952) evidence and concepts of autonomic response specificity.

Since there are theoretical advantages and disadvantages to each method in the present experiment, analyses were performed with all three methods. Occasionally the data are reported with respect to one physiological method of measurement, while in other instances another method has been presented because more striking relationships were obtained with it.

In our earlier work (Vogel et al., 1958), the most meaningful results were obtained with the first method, that of Lacey. In the present study we obtained intercorrelations between the three physiological measures, with an *N* of 40, separately for the ego-involvement and stressor conditions. Under the ego-involvement condition, Lacey's method of using the most reactive channel (ignoring conditions) correlates $+.90$ with the method of using the highest channel for each condition, and $+.57$ with the simple average of all channels. The measure based on the highest channel for each condition correlates $+.73$ with the average for all channels. Under the

stressor condition, the pattern is similar, the respective correlations being $+.96$, $+.65$, and $+.73$. All of these relationships are highly significant and substantial in magnitude, and in the context of confusion in the literature over which type of approach to employ, it is possible to argue for a certain amount of interchangeability between them. Most commonly, the direction of relationships with one method is the same as that found with another, although, at times, the significance level changes.

Tasks

In an earlier related study (Vogel et al., 1958), the McKinney Reporting Task (McKinney et al., 1951) had been employed. It consists of a long series of circles containing three symbols in various quantities in each. The *S*'s task is to count as quickly and accurately as possible the number of symbols of each kind that each circle contains. This task was also employed in the present study.

In addition to this rather simple perceptual-motor task, a second task was also employed in which the *S* had to unscramble a set of scrambled words as quickly and as accurately as possible. In order to examine performance on these two tasks under comparable psychological conditions, *Ss* in each experimental condition worked on them alternately, with four one-minute periods of each.

Experimental Conditions and Procedure

The entire experiment took place in two sessions. The first session comprised the pre-selection testing; the experimental procedure proper took place two months later. *Ss* were tested individually over a two-week period. The *Ss* selected from the initial testing were sent letters requesting their cooperation and indicating that they would receive pay at the rate of \$1 an hour. All of the *Ss* solicited in this way appeared for the experiment. The first condition of the experimental session may be identified as the practice period (following accustoming the *S* to the physiological apparatus) in which the *S* was given first the McKinney Reporting Test and the scrambled words task in alternating fashion, with the two different types of task in order that he might get used to them. Through this point the treatment accorded all *Ss* was precisely the same.

Each of the four pre-selected groups had been subdivided into two matched groups which from this point on were given different treatments: on the one hand, an affiliation stressor condition, and on the other hand, an achievement stressor condition. In other words, half of the achievement oriented *Ss* were exposed to an achievement stressor and half to an affiliation stressor. Similarly, half of the affiliation oriented group were exposed to an achievement stressor and half to an affiliation stressor.

Following the practice condition, the ego-involvement condition was introduced. For the achievement stressor treatment, *Ss* were told that the testing, for which they had been selected, measured their capacity for successful academic and occupational achievement. To lend credence to these instructions the experimenter ostensibly read them from a copy of the *Journal of Educational Psychology*. The instructions were as follows:

The sole purpose of the Clark University study is to measure degree of academic and intellectual ability. It has already been shown in previous research that persons who are outstanding in general intellectual attainment, as well as those who are outstanding in their college studies, do superior work on both parts of the test. Further work is now being carried out with the test. For instance, psychologists have found they can best discover which preparatory students will become academic successors in college by checking both physiological reactivity during the test, as well as performance on both the scrambled words and reporting task sections of the test.

The affiliation stressor treatment, on the other hand, included the following instructions which emphasized capacities for warm affiliative relations with other people. These instructions were pasted into, and ostensibly read to the *S* from, a copy of the *Journal of Personality*:

The sole purpose of the Clark University study is to determine which persons have the qualities which make warm, friendly personalities. The measures are not related to intellectual or social attainment, but are concerned with identifying those attributes which make a person warm and friendly in interpersonal relations. It has already been shown during this research that people who have these particular qualities do superior work on both parts of the test. For instance, psychologists have found that they can best discover which preparatory students have these qualities of warmth and friendliness by checking the *S*'s physiological reaction during the test, as well as the *S*'s performance on both scrambled words and reporting task sections of the test.

At this point certain aspects of the tasks were changed. The symbols on the McKinney Reporting Test were suddenly made to assume relevance. For the affiliation stressor treatment, the symbols represented "friend," "acquaintance," and "enemy"; for the achievement stressor treatment they represented "honors," "pass," "flunk." With respect to the scrambled words task, in the practice session both groups had received the same innocuous scrambled words such as "truck," "window," "breeze," etc. Now in the ego-involvement condition, all the words were need related. The achievement stressor treatment involved such words as "fail," "poorest," "teacher," "school"; the affiliation stressor treatment included such words as "mother," "company," "lonely." The words given the two groups were matched for frequency in terms of the Thorndike-Lorge List (1944), for word length, and method of scrambling. During the ego-involvement condition, Ss performed the tasks just as they had done during the practice session with eight minutes of work, one minute alternately on each task, beginning with the McKinney Reporting Test.

Following the ego-involvement session, the experimenter examined each S's performance and initiated the failure-stressor condition by informing him that he had failed the test very badly. The achievement stressor and affiliation stressor treatments involved different threatening statements, related to the need state being assaulted. For the achievement-stressor treatment, Ss were told that their prospects for college and future successes were poor, and it was suggested that it might be worth trying again. For the affiliation stressor treatment, the person was identified as a person who lacks capacity for warm, friendly relationships compared with others. It was admitted that he might have a need for friendly relations but he was just not warm and friendly as a person. It was implied that all the information obtained by the experimenter supported this conclusion. Following these statements, the task performance was resumed exactly as in the ego-involvement condition, except that every two minutes the experimenter, examining the S's work, extemporaneously criticized its poor quality.

RESULTS

In our presentation of the results, we shall separately examine the data pertinent to the three issues cited at the outset. First, we consider the proposition that stress depends upon the relationship between the stressor condition (achievement- or affiliation-stressors) and the predominant intrinsic motive of the S.

The Role of Intrinsic Motivation in Determining Stress Reactions

The empirical consequence of our proposition that stress depends upon the relevance of the stressor condition to the motive pattern can be examined by comparing the Ss who are achievement oriented with those who are affiliation oriented in respect to physiological arousal under the two types of ego-orienting conditions. Table 2 presents these data. We have presented the physiological reactivity as determined by the various subject-condition interactions, in terms of each of the three procedures for combining physiological measures discussed in the procedure section.

While the trends for each of these procedures are similar, there are significant interactions between type of motive orienting condition and motive pattern for the first two, but not for

TABLE 2
PHYSIOLOGICAL AROUSAL AS A FUNCTION OF TYPE OF
INTRINSIC MOTIVATION AND THE NATURE OF THE
EXPERIMENTAL CONDITIONS

Groups and Conditions	Highest Channel ^a (Lacey)	Highest Channel for Each Condition ^b	Average for All Measures ^c
Achievement motivation			
achievement oriented conditions	60.7	62.3	54.1
affiliation oriented conditions	57.2	59.3	50.6
Affiliation motivation			
achievement oriented conditions	58.6	60.1	51.9
affiliation oriented conditions	63.1	64.0	52.8

Note.—There is a significant main effect source of variation when the ego-involvement condition is compared with the stressor condition. These figures are derived by summing across these two conditions because the patterns shown here are comparable within the ego-involvement and stressor conditions and tend to be significant in each case. This breakdown is simpler and more to the point in examining the role of motivation in determining arousal under stressor conditions.

^a $F = 5.58$; significant at less than .05 level.

^b $F = 5.00$; significant at less than .05 level.

^c $F = 2.88$; not significant.

the third. In other words, when the achievement oriented group is presented with achievement oriented conditions (conditions that pose a relevant threat to the central motive state of the *S*), there is a high degree of autonomic reactivity. On the other hand, when the achievement oriented group is exposed to a nonrelevant threat (the affiliation oriented conditions), there is a low degree of reactivity. Similarly, when the affiliation oriented group is exposed to affiliation orientation, the physiological reaction is greater than when this group is exposed to an achievement-orienting condition. A further analysis of these results in which the ego-involvement condition is separated from the stressor condition produces no significant interactions. The pattern is the same for both experimental conditions, and it occurs regardless of the method of combining physiological indices, although it is more striking and statistically significant with some methods than with others.

The Relation of the Stress State to Task Performance

There does not exist a simple relationship between the stress state and performance, a finding that was expected in the light of the literature on stress and performance and on the basis of previous research by Vogel et al. (1958), where it was found that performance depended upon the interplay of motivation and somatic reactivity. Moreover, we have hypoth-

esized different effects of the stress state depending upon whether the task was essentially perceptual-motor or conceptual in character. In the present study, no significant relationships were found between performance and the motivation variable alone, or between performance and the relevance of the stressor conditions to intrinsic motivation. However, as has been previously found in similar research (Vogel et al., 1958), significant effects in performance were obtained when stress relevance and physiological reactivity were studied in interaction.

If we combine performance on both the McKinney Reporting Test and the Scrambled Words Task in terms of percentage of items correct during the stressor condition, we find that when the *S* is exposed to a stressor condition relevant to his intrinsic motivation, his performance is good if he also responds to the stressor condition with a *high* degree of physiological reactivity. However, under stress relevant conditions and with a *low* degree of physiological reactivity, performance is likely to be poor. On the other hand, when the stressor is irrelevant to the predominant intrinsic motivation and there is a high degree of physiological reactivity, performance is less adequate than when there is a low degree of physiological reactivity. These data are presented in Table 3 with the McKinney Reporting Test and the Scrambled Words Task shown separately, as well as combined. This finding replicates one previously reported (Vogel et al., 1958) with the McKinney Reporting Test where it was argued that *Ss* who were aroused even when it is inappropriate or, conversely, unaroused when they should be aroused, showed inadequate performance functioning. It should be noted that in the over-all analysis of variance, no significant effects were contributed by type of stressor or type of motivation by themselves. Therefore it was possible to combine them in the form of motivation-relevance categories.

An interesting relationship between stress relevance and performance as a function of type of task appears in Table 4. These data deal with our hypothesis that stress will facilitate certain sensory-motor functions but interfere with conceptual ones. It will be seen that a motivation relevant stressor is associated

TABLE 3
PERFORMANCE AS AN INTERACTIVE FUNCTION OF STRESS RELEVANCE AND PHYSIOLOGICAL REACTIVITY UNDER THE STRESSOR CONDITION

Group	N	Percent Items Correct		
		McKinney Reporting Test	Scrambled Words	Both Tasks Combined*
Motivation relevant stressor				
High physiological reactivity	10	50.56	61.34	55.95
Low physiological reactivity	10	39.46	48.51	43.98
Motivation irrelevant stressor				
High physiological reactivity	10	49.58	51.06	50.32
Low physiological reactivity	10	45.34	56.98	51.16

Note.—The index of physiological reactivity here is based on averaging the measures in each channel and making a median split. The other methods produce only comparable trends.

* *F* for interaction = 6.34; significant at below .05 level.

TABLE 4
PERFORMANCE AS A FUNCTION OF STRESS RELEVANCE
AND TYPE OF TASK*

Group	N	Ego-Involve- ment		Stressor	
		McKinney	Scrambled Words	McKinney	Scrambled Words
Motivation relevant stressor	20	55.06	45.98	57.94	53.49
Motivation irrelevant stressor	20	52.25	55.70	54.10	56.31

Note.—Performance measured by number items attempted. Similar trends are found for percentage items correct.

* Interaction F (motivation \times tasks \times conditions) is significant below .05 level.

^a Essentially the same pattern of results is found if high and low physiological reactivity or arousal is substituted for the motivation relevance categories.

with low performance output on the conceptual task (scrambled words) compared with a motivation irrelevant stressor. The reverse pattern, however, holds for the perceptual-motor task (McKinney Reporting Test). A motivationally relevant stressor is associated with higher output than a motivationally irrelevant stressor. It should also be noted parenthetically that the same kind of relationship is found with degree of physiological reactivity as the means of S classification; physiological arousal and motivational relevance have been shown to be related in Table 2. Moreover, type of stressor and type of motivation separately produced no significant effects, making the combined categories of motivation relevance appropriate.

Past Success and Failure and the Stress Reaction

At this point it should be recalled that both the affiliation oriented and achievement oriented S s were divided into those considered successful in gratifying these needs and those who were unsuccessful. By means of the factorial design, it was possible to study the contribution of this variable to the arousal of a state of stress. Table 5 demonstrates that degree of success in fulfilling one's predominant motivations is a significant factor in determining the degree of affective arousal. Successful people are more readily aroused in the ego-involvement condition (minimal threat) than unsuccessful people, but, in turn, their increase in physiological arousal under the stressor condition (strong threat) is less than for the

TABLE 5
SUCCESSFULNESS AT MOTIVE GRATIFICATION IN
RELATION TO PHYSIOLOGICAL AROUSAL

Group	N	Physiological Arousal ^a	
		Ego-involve- ment	Stressor
Successful	20	55.43	64.28
Unsuccessful	20	52.17	67.68

^a The method of combining physiological measures here is the use of the highest channel (Lacey's method). With this method the interaction $F = 4.49$; significant below the .05 level. With the use of the highest channel for each condition, the F is 3.43 approaching significance. The data are comparable in direction with a simple averaging of physiological measures, but do not reach statistical significance.

unsuccessful group. Further analysis has shown that there is no interaction between success and stress relevance, or success and the nature of the motive pattern. Moreover, there was no significant effect of past degree of success on any of the performance measures.

SUMMARY

Three issues were explored experimentally in this study. One dealt with the proposition that the arousal of a state of stress depends upon intrinsic motive states in the individual as well as the nature of the stressor conditions. A second was concerned with the relations between a stress state and task performance. The proposal was made that stress would tend to improve sensory-motor output and impair conceptual performance on the basis of different amounts of interference. The third involved the importance of the S 's past history of success or failure in a particular motivational direction in determining whether a stress state would be aroused.

One hundred and eighty-five preparatory school students were used as a general subject pool and were given a battery of tests and procedures in order to select two groups differing in predominant motive pattern. Twenty S s were selected as strongly oriented toward achievement as opposed to affiliation, and 20 with the reverse pattern, representing as ideal cases as it was possible to select to contrast the two types of predominant motivations. Subjects were additionally divided into those who were successful and those who were unsuccessful within their motive sphere, and arranged in a factorial design. Half of the achievement oriented S s were exposed to an achievement

oriented type of stressor condition and half to an affiliation stressor condition. The affiliation oriented group was similarly divided and given the two stressor treatments. The arousal of stress was studied by means of autonomic reactions, including pulse, blood pressure, and GSR.

The results clearly support the view that the arousal of stress depends upon the relationship between the motive pattern and the type of stressor condition. The relative effect of stress on a sensory-motor task was facilitative, while on a conceptual task it produced impairment. Successful subjects were more easily aroused by minimally threatening conditions than unsuccessful subjects, and less disturbed by more severe stressor conditions.

In general, the research suggests the importance in stress research and theory, of such factors as intrinsic motivation, the type of task, and the S's past history of success or failure within the relevant motivational sphere.

REFERENCES

- BROVERMAN, D., & LAZARUS, R. S. Individual differences in task performance under conditions of cognitive interference. *J. Pers.*, 1958, **26**, 94-105.
- IVERSON, M. A., & REUDER, M. E. Ego-involvement as an experimental variable. *Psychol. Rep.*, 1956.
- LACEY, J. I. Autonomic response specificity and Rorschach color responses. *Psychosom. Med.*, 1952, **14**, 256-260.
- LACEY, J. I. The evaluation of autonomic responses: Toward a general solution. *Ann. N. Y. Acad. Sci.*, 1956, **67**, 123-164.
- LAZARUS, R. S., & BAKER, R. W. Personality and psychological stress—A theoretical and methodological framework. *Psychol. Newsltr.*, 1956, **8**, 21-32.
- LAZARUS, R. S., & BAKER, R. W. Motivation and personality in psychological stress. *Psychol. Newsltr.*, 1957, **8**, 159-193.
- LAZARUS, R. S., BAKER, R. W., BROVERMAN, D. M., & MAYER, J. Personality and psychological stress. *J. Pers.*, 1957, **25**, 559-577.
- LAZARUS, R. S., DEESE, J., & OSLER, S. J. The effects of psychological stress upon performance. *Psychol. Bull.*, 1952, **49**, 293-317.
- MCCLELLAND, D. C. *Personality*. New York: Dryden, 1951.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
- McKINNEY, F., STROTHER, G. B., HINES, R. R., & ALLEE, R. A. Experimental frustration in a group test situation. *J. abnorm. soc. Psychol.*, 1951, **46**, 316-323.
- PHILLIPS, L. & COWITZ, B. Social attainment and reactions to stress. *J. Pers.*, 1953, **2**, 270-283.
- SARASON, S. B., & MANDLER, D. Some correlates of test anxiety. *J. abnorm. soc. Psychol.*, 1952, **47**, 561-565.
- THORNDIKE, E. L., & LORGE, I. *The teachers' word book of 30,000 words*. New York: Columbia Univer. Press, 1944.
- VOGEL, W., BAKER, R. W., & LAZARUS, R. S. The role of motivation in psychological stress. *J. abnorm. soc. Psychol.*, 1958, **56**, 105-112.

Received December 9, 1957.

PERCEPTUAL SETS IN KORSAKOFF'S PSYCHOSIS¹

GEORGE A. TALLAND AND ALICEMARIE MILLER

Harvard Medical School and Massachusetts General Hospital

KORSAKOFF'S psychosis is characterized by a severe disturbance of the capacity to remember and to learn, and only to a much smaller extent by other cognitive or affective disorders. Understandably enough, few psychological studies of this disease have paid attention to perceptual processes, or even considered them relevant to the principal problem of Korsakoff patients, which is their virtual inability to learn or to recall anything they have experienced since the onset of their illness and a retrograde amnesia of varying degrees. The earliest psychological research on Korsakoff's psychosis followed a simple associationist model. However, in the works of Pick (1915) and Gruenthal (1923), the ideas of the Würzburg school, in particular those of Ach (1905) and Selz (1913), soon made an impact on the study of the amnesic syndrome. This new departure in the study of the amnesic syndrome took the direction of explaining the Korsakoff patient's cognitive disorder in terms of his more complex thought processes. The outstanding contribution of the Würzburg school to psychological theory, however, proved to be not so much its analysis of thought processes as the concept of *Einstellung*, translated in various contexts as mental set, disposition, attitude, or orientation. This concept certainly influenced Gruenthal's observation about the Korsakoff patient's inability to shift from one train of thought to another. Whether an attitudinal inflexibility thus defined characterized behavior in Korsakoff's psychosis more generally could have been tested on perceptual tasks. While Gruenthal did not extend his observations beyond the range of thought processes, his critics, Buerger-Prinz and Kaila (1930), working under the influence of gestalt theory, did somewhat bridge the gap between perception and thinking in the study of Korsakoff's psychosis. Thus, these authors complemented rather than re-

futed Gruenthal's theory, as they thought they had done, by pointing out that the Korsakoff patient's inability is not one of changing but that of assuming certain attitudes or dispositions, and more particularly an incapacity to adopt an orientation towards the total situation. However, such an overwhelming preoccupation with details as would prevent the perception of the whole is not the same thing as an incapacity to shift one's orientation to the whole from the details after an orientation to the latter has been established. All that Buerger-Prinz and Kaila's experiment demonstrated was that Korsakoff patients were unable to shift their orientation, thus drawing attention to the significance of attitude in the analysis of the amnesic syndrome. This evaluation of their contribution is in agreement with Lidz's (1942). Similarly, the passivity characteristic of the disease implies a deficiency in that affective component which, according to Bartlett (1932), is involved in the constructive and selective processes of perceiving and remembering.

That perception is a joint product of sensory input and such selective and integrating dispositions as expectancy, familiarity, and interest, is a view now generally endorsed and supported by considerable experimental evidence (Allport, 1955). Failure or deficiency in perceptual performance, therefore, can be caused by the defective operation of these dispositions, as well as by impairment of the sensory apparatus. Experimental techniques have been developed by which the influence of these two sets of factors can be controlled and varied. One of these methods is that of presenting stimuli at the threshold level of accurate perception, testing the operation of the selective or structuring influences by their effect in raising the sensory input above the limen.

Our interest was in ascertaining to what extent dispositions of this type would help Korsakoff patients in accurate perception. Using expectancy as a perceptual set, the fol-

¹ This research was supported in part by a Research Grant, M-767(c), from the National Institute of Mental Health, United States Public Health Service.

lowing hypotheses were subjected to experimental testing: As compared with the control group,

(a) For Korsakoff patients, an expectancy of the order of a single set is less effective, as tested by the accurate perception of congruent stimuli and the systematic distortion of incongruent stimuli;

(b) An expectancy of the order of a multiple set is further reduced in its effect with Korsakoff patients, and more particularly they are able to utilize only one out of two or more sets;

(c) Korsakoff patients are less effective in developing expectancies;

(d) Korsakoff patients are less effective in changing sets.

METHOD

Subjects. The experimental sample consisted of 21 patients from the Boston State Hospital, all of whom had a history of alcoholism and were diagnosed by the neurologist members of our research team as having Korsakoff's psychosis. They exemplified the chronic phase of this disease, having been hospitalized for periods ranging from about one year to over twelve years, and before that had been treated in other hospitals for the acute symptoms of Wernicke's encephalopathy. They formed about three-fifths of a larger sample from which they were selected for being relatively least damaged in their cognitive functioning. Most of the remainder were too severely deteriorated to follow even simple instructions, and therefore their test performance would in no way have added to our understanding of the specific characteristics of Korsakoff's psychosis. In age, the experimental *Ss* ranged from 31 to 74 (mean 54.0; *SD* 9.77); in intelligence from 83 to 117 (mean 102.1; *SD* 9.26) as measured by three verbal subtests (comprehension, vocabulary, similarities) of the Wechsler Bellevue battery (Wechsler, 1944). Twenty members of this original group participated in Experiment II.

Twelve members of the control group were outpatients in the Alcoholic Clinic of the Massachusetts General Hospital, six were patients on the neurological wards of the Lemuel Shattuck Hospital. The former control *Ss* were chosen for having a history of alcoholism in common with the experimental sample but, like these, no alcohol in their system at the time of testing; abstinence was a condition of treatment at the psychotherapy clinic. The latter control *Ss* were selected for sharing with the Korsakoff psychotics the experience of prolonged hospitalization. They had all been on the wards for at least six months, but they had neither a history of alcoholism nor any known brain damage. In age they ranged from 35 to 60, (mean = 46.4; *SD* = 7.55). In IQ they also represented the average sector with a few exceptions of superior intelligence among the alcoholic outpatients; based on scores of the three subtests (CVS), they ranged from 92 to 140 with a mean of 114.1 (*SD* = 10.75).

Procedure. The experimental material consisted of lists of words recorded on tape. These words were presented by a tape recorder for vocal reproduction at or near the

threshold level of intelligibility. Several of the lists consisted of words drawn entirely from one conceptual class, e.g., all animal names; others combined two classes; yet others mixed classified and unclassified words. Set was defined as the expectancy of words belonging to one or another category, and its operation was tested by accurate reproduction of the more or less ambiguous stimuli. Another observable effect of mental sets was an incorrect reproduction of the stimulus, but one congruent with an expectancy. Instances of this type of error were recorded and are referred to as *assimilation*.

Expectancies can either be established explicitly or implicitly. An *explicit* set was formed for *S* by instruction or warning, e.g., that he would hear names of animals. *Implicit* set, in this experiment, means that *S* was not advised about the selection of the words he would hear but that an expectancy was induced in him by means of a pre-list of words drawn from the same categories as those of the test list. The establishment of an appropriate set could be hindered as well as facilitated, e.g., when *S* was advised that he would hear animal names, and was indeed presented with a list of these, but at some point the auditory material suddenly changed to names of parts of the body. This situation is referred to as *counter set*. Sets were *single*, e.g., when the expectancy was for only one class of words, or *multiple*, when names drawn from two or more categories were expected to occur. Test lists in this experiment provided either *complete confirmation* of the set, i.e., every item was congruent with expectancy, or *partial confirmation*, when only some of the items were congruent and others were not.

The following types of set were established: single set, both explicit and implicit, with complete and also with partial confirmation; double set, explicit, with complete and with partial confirmation; single counter set, following complete confirmation. Each set was confirmed by a list of 12 words, and the same number of mixed words was used for "padding" in partial confirmation. Test lists therefore varied in length; they consisted of 12 items for an explicit single set with complete confirmation and another 12 for a counter set. Thus, 24 words were needed to test a double set with complete confirmation or a single set with partial confirmation; 36 for a double set with partial confirmation. An implicit set was induced by a pre-list of 12 words, which was followed without break by the 12 words of the test series. Lists of 24 words or longer were so constructed that not more than two names drawn from the same category or two mixed words would follow in a run.

Expectancies were established for one or two of the following classes of words: animals, food, parts of the body, garments, birds, colors. Alternate lists of 12 were compiled from names representing the first three of these categories, and there were also three lists of mixed words, no two of which within an array of 12 were drawn from the same conceptual class. Since frequency count in print did not prove a reliable measure of aural intelligibility, these mixed words were selected after a pilot test, so that each list of 12 closely matched the associated classified list in intelligibility. These lists

are referred to by the initial letter of the generic term, e.g., A for animals, F for food, P, G, B, C for the other categories, and X for mixed words. Suffixes distinguish alternate lists within one class.

The study consisted of two experiments. The first tested the effect of different types of set on Korsakoff patients against their effect on the control Ss. Since the various test lists had not been equated for basic intelligibility, a second experiment was designed to test the relative effect of one type of set as against another, and of partial as compared with complete confirmation. This second experiment was conducted on a different occasion and was limited to the experimental group.

Experiment I consisted of the following seven tests: implicit single test, partial confirmation (F_1X_1); implicit single set, complete confirmation (G); explicit double set, complete confirmation (A_1P_1); explicit single set, partial confirmation (F_2X_2); explicit double set, partial confirmation (BCX_3); explicit single set, followed by counter set (P_2-A_2).

In Experiment II, (F_2X_2) was presented after an implicit, (F_1X_1), after an explicit set had been induced for food names. Half the experimental group was given (A_1), the other half (A_2) with an implicit set for animal names. The alternate animal list always followed with an explicit set. Last came (P_2) to test the effect of counter set. The same tape was used for both experiments with instructions appropriately altered.

S was seated at a distance of about 5 ft. from the Revere tape recorder and was instructed to listen and to do exactly as he would be told over the instrument. Instructions were also recorded on the tape, but at a considerably higher level of loudness than the test lists. The first instruction read: "You will hear a list of words spoken one at a time. Listen carefully, and repeat out loud each word after you hear it. If you are not sure what word you heard, repeat what you think you heard. The words you will hear will be just loud enough for you to hear, so you may not understand every word you hear. Don't let that discourage you; repeat out loud every word as it *sounds* to you. Don't make any other

remarks, just repeat each word once." Then followed a pilot list of 24 words, which included a sequence of 12 furniture names. Responses to this list were used to establish the S's threshold level. The volume and tone dials of the instrument were set so that he made as many correct responses as errors by incorrect or omitted reproduction.

The test proper began with an implicit set, for which the instruction was: "Now you will hear another list of words. Again, repeat every word just as it sounded to you." Subsequent instructions did not reiterate this caution, nor was there any need for E to remind S what his task was. Implicit set, on the second occasion, was induced by simply announcing another list of words, explicit set by the following type of instruction: "Next you will hear names of food, such as asparagus or tea," or "This time you will hear names of birds and names of colors, such as thrush or purple."

The silent interval on the tape after each word was long enough for S to repeat four or five times what he had heard. S's responses were recorded under four headings: correct; none; incorrect and meaningless words; assimilation. Examples of the latter would be *tea* for *cheese*, or *savory* for *favorite*, when set was established for food names.

RESULTS

Judged by the results set out in Table 1, Korsakoff patients are evidently capable of utilizing expectancies, since the mixed words were not inherently less intelligible than the classified names for which sets were in operation. The experimental group's mean score is below the control sample's in every instance, but significantly so only in the tests of implicit set with complete confirmation ($t = 3.26$, $p < .01$) and explicit set with partial confirmation ($t = 2.36$, $p < .05$). The facilitative effect

TABLE 1
MEAN CORRECT REPORT FROM LISTS OF 12 WORDS; ASSIMILATION IN FREQUENCY AND AS PERCENTAGE OF TOTAL ERROR

Set	Confirmation	List	Assimilation					
			Mean correct report		Frequency		Percentage of errors	
			Korsakoff	Control	Kors.	Contr.	Kors.	Contr.
Single								
Explicit	Complete	P_2	8.95	9.94				
Explicit	Partial	F_2	5.57	7.72	18	19	28	51
Implicit	Complete	G	6.57	8.50	14	10	10	14
Implicit	Partial	F_1	7.71	8.50	31	15	27	24
Counter	Complete	A_2	6.19	6.83	18	9	20	14
Double					15	15	12	16
Explicit	Complete	A_1	5.19	7.95				
		P_1	6.95	8.83	38	24	27	33
Explicit	Partial	B	5.81	9.72	24	17	16	30
		C	5.90	8.56	7	3	5	7
Mixed Word Lists		X_1	2.71	4.06	3	7	2	11
		X_2	1.67	1.94	29	29	15	20
		X_3	2.52	2.39	26	23	12	13
					9	16	5	9

of double sets, however, is significantly lower for Korsakoff psychotics. In fact, even though advised to expect words representing either of two classes, they seemed to function with only one set—either for animals or parts of the body, either for birds or colors. Mean correct responses to either of the matched classes did not differ significantly whether confirmation was complete or partial or in either the experimental or the control group. However, they did differ significantly when the relative effectiveness of the two joint sets, irrespective of which happened to be stronger, was taken into account. This can be seen in Table 2, which presents mean correct responses in accordance with the more and the less effective of two simultaneous sets for each *S* without regard to the fact that the stronger set was for birds in one *S*, for colors in another. Clearly, the difference between the two samples is largely due to the operation of the minor or less effective set. Differences between recognition score in accordance with the major and minor set were calculated for individual *Ss*, and the difference between the two sample means was tested for significance. In the case of double set with complete confirmation, the means were 4.62 and 1.89. Because the variances of the two distributions were not homogeneous, the test was based on ranked data, using Mosteller and Bush's (1954) formula for *T*. The *z* score obtained was 3.51, which means that even this comparatively insensitive test shows a difference between the two samples, significant at the .01 level. Correct responses with double set and partial confirmation provided three arrays of differential scores, means of which are set out in Table 3. Student's *t* was applicable to test the significance of the difference between each of the means paired, and all three *t* ratios are significant with *df* = 37. With a single set in operation, the difference between

TABLE 3
MEAN DIFFERENCE BETWEEN CORRECT RESPONSE IN ACCORDANCE WITH MAJOR SET, MINOR SET AND MIXED WORDS; AND TEST OF DIFFERENCE BETWEEN SAMPLE MEANS

	Korsakoff	Control	<i>t</i> ratio	<i>p</i>
Major-Minor set.....	3.29	1.39	2.77	<.01
Major set-Mixed words.....	4.95	7.47	3.31	<.01
Minor set-Mixed words.....	1.70	6.00	6.61	<.01

correct report of classified and mixed words was larger in the Korsakoff group when set had been induced implicitly; in the control group, this was so when set had been established explicitly. In neither instance did the two groups differ significantly.

In terms of assimilation, there were no noteworthy differences between the Korsakoff and control groups, no matter whether these errors are compared by incidence or as a percentage of the total error. The few examples of significant differences occurred with a double set. Even the small total of 26 instances of assimilation by the control group from list BCX₃ is significantly larger in proportion to their total error score than the Korsakoff group's 19 (chi square = 8.64, *p* < .01). Yet more closely in line with the previous finding was the fact that out of the latter group's 62 assimilations from list A₁P₁, 45 were in accordance with the major set, 13 with the minor set, and 4 occurred when neither of the two sets was dominant. The corresponding breakdown for the control group was 16-18-7. This trend was less pronounced when the double set was only partially confirmed; only 11 out of the Korsakoff group's 19 instances of assimilation were congruent with a dominant set. On the other hand, the reverse tendency operated in the control group, where 5 examples of assimilation followed the major and 13 the minor set. The remaining 8 occurred when neither set was dominant.

Contrary to their instruction to repeat each word as it sounded to them, *Ss* of both groups made more errors by failure to respond than by incorrect response, the latter including assimilation. In the experimental and control groups, the ratios of these two types of error were 1.41 and 1.13. The Korsakoff patients' stronger tendency to resort to silence when in doubt is apparent also from their individual

TABLE 2

MEAN CORRECT RESPONSE IN ACCORDANCE WITH MAJOR AND MINOR OF TWO CONCURRENT SETS

	Double Set Complete Confirmation		Double Set Partial Confirmation	
	Major set	Minor set	Major set	Minor set
Korsakoff.....	8.38	3.76	7.50	4.21
Control.....	9.33	7.45	9.83	8.44

TABLE 4
MEAN CORRECT RESPONSE IN EXPERIMENT II

Set	Confirmation	List	Mean Score
Explicit single	Partial	F ₁	7.95
Implicit single	Partial	F ₂	5.35
Explicit single	Complete	{ A ₁	8.20
		{ A ₂	8.00
Implicit single	Complete	{ A ₁	6.40
		{ A ₂	6.10
Counter single	Complete	{ P ₂	6.00
Mixed word lists		{ X ₁	1.65
		{ X ₂	2.40

ratio scores. While exactly half the control Ss made more errors by incorrect call than by no response, only five of the experimental group did so. However, the difference between the mean ratios of the two groups, 2.31 and 1.93, is not statistically significant.

The results of Experiment II, as shown in Table 4, resolve several questions and one paradox raised by Experiment I by demonstrating the differential basic intelligibility of the test lists used in it. Whether confirmation is complete or partial, an implicit set does not facilitate correct perception more effectively than an explicit set, and for Korsakoff psychotics no more than for others. With means of 6.63 and 6.56, calculated over the two experiments, the two types of set do not differ in their facilitative effect, whereas lists F₁ and F₂ do with means of 7.38 and 5.46 ($t = 4.23$, $p < .01$). With complete confirmation, on the other hand, the facilitative effect of an explicit set is significantly greater than that of an implicit set. This finding of Experiment I was confirmed by testing the difference between mean correct responses to lists A₁ and A₂ under the two experimental conditions ($t = 4.20$, $p < .01$).

Scores on list P₂ show that a counter set, as would be expected, had a smaller facilitative effect on correct recognition than an explicit set. However, a counter set seemed to exert no inhibitory effect on Korsakoff patients, whose correct responses to list A₂ were just a little less with an implicit set than it had been in Experiment I with a counter set. The closest test available for comparing the effects of complete and partial confirmation on accurate

perception with an explicit set in operation is that provided by list A₁ in the two experiments. For the 14 experimental Ss with whom animals were the major set, Test A₁P₁ of Experiment I can be regarded as one of an explicit set for this class of words, followed by partial confirmation. Their mean score on A was 8.16, as compared with mean correct report of 8.20 from the same list in Experiment II. In both Experiments, a larger number of mixed words was reported correctly when the set was established implicitly than with an explicit set, and the difference is significant ($t = 5.37$, $p < .01$).

DISCUSSION

The experimental results have confirmed the first three hypotheses, but not the fourth, and the evidence on the first is not conclusive. In so far as experimentally induced expectancies are typical of perceptual sets or attitudes in general, Korsakoff patients as a class differ from others in their capacity to develop and exercise situationally appropriate dispositions. It has also been demonstrated that these patients are not entirely incapable of establishing or utilizing such selective processes as perceptual sets, that they do not select randomly from the environmental stimuli, and that their selective operations are not qualitatively different from the normal, but are narrowly limited.

The Korsakoff patient can maintain and effectively use a single set expressly indicated to him. He is, however, at a disadvantage when he has to create such a set for himself from cues provided by the situation. Experiments (Talland, 1958) with the same type of test material as was used here have shown that normal people gain as effective an expectancy from a pre-list of six words as from an explicit instruction for a set, and that a counter set seriously interferes with their accurate perception. Our finding that the effect of a counter set on Korsakoff patients is much the same as that of an implicit set suggests also that they are no more rigidly committed to an expectancy they have adopted than are controls; they do not persevere.

In this study, we have not explored the limits within which the Korsakoff patient can effectively maintain a firmly established set, i.e., the amount or spacing of disconfirmation

that would terminate it or the possibility and necessary conditions of its re-arousal. Our findings, however, concur with other observations which suggest that the set adopted will persist as long as it is regularly confirmed and will be irrevocably relinquished in the prolonged absence of confirmation. Assimilation to the previously confirmed set occurs with considerable frequency immediately after the introduction of a counter set and hardly at all later, although the errors are fairly evenly distributed along the list. Yet instances of complete perservation occur even in this psychosis. For example, one *S* either assimilated to parts of the body or kept silent throughout the entire list of animal names presented under a counter set. More common were the examples of another kind of perseveration. Food names tended to crop up in the Korsakoff patients' report with a regularity which bore no relation to the recency of an appropriate set for this class of words. No doubt, this set persevered from an extra-experimental situation of these hospitalized *Ss* who, we had observed, still maintained some interest in and expectancy of their meals in spite of their comparative apathy and disorientation in time.

Much the most significant difference between the Korsakoff and the control groups appeared in this study when they had to exercise more than one set in order to maximize correct perception. The psychotics seem incapable of carrying two or more sets, switching from one to the other all the time. This finding fits into Gruenthal's theory and extends it to perceptual processes; it also agrees with Buerger-Prinz and Kaila's. The complementary finding of the effect of counter sets on Korsakoff patients, however, cannot be easily reconciled with Gruenthal's argument. The Korsakoff group's difficulty is not so much in abandoning one orientation for another, but in adopting a potential readiness in addition to an actual one. In gestalt terms, the two lists of a double set have initially the same properties of good form, so that as soon as an attitude appropriate to a figure is directed toward one, the other becomes the ground. With this reversible perspective, normals can change figure and ground with ease; Korsakoff patients cannot perceive the ground as figure. Designs with reversible perspective are arti-

ficial and uncommon, but situations requiring alternative perceptual sets are not. The Korsakoff patient's inability to carry more than a single expectancy may, therefore, be more germane to his characteristic cognitive disorder than his concern with details, which he shares with other clinical classes. It may be merely a function of his memory defect, in which case it has no explanatory value, though our study still serves a purpose by demonstrating that this defect is not so complete as to prevent entirely the formation and maintenance of expectancies. On the other hand, one can regard the Korsakoff patient's limited capacity to develop new or to carry alternate sets as symptomatic of his generally restricted, passive response to the environment, rather than an active perception of it. If those centrally initiated processes called attitudes or mental sets, which exercise a selective and integrative function in perception, also determine the reconstructive act of remembering, our findings converge on the major cognitive symptoms of Korsakoff psychosis. Patients suffering from this disease illustrate, by providing a negative instance, Bartlett's difficult model of "the organism's capacity to turn round upon its own 'schemata'" (Bartlett, 1932, p. 213). More likely yet, their disability is not so much in recreating as in first establishing the attitudes which integrate experiences or behavior into an organized system. Their *schemata* are deficient; they have nothing to turn back on, except their premorbid *schemata*.

SUMMARY

On the suggestion that the cognitive disorder characteristic of Korsakoff's psychosis may be, at least in part, attributed to the patients' incapacity to adopt and maintain attitudes appropriate to the situation or task, a sample of 21 was tested for the effect different types of set would have on their perception of auditory stimuli presented near the threshold level of intelligibility. The control group was composed in part of alcoholic outpatients in a psychiatric clinic, who were not known to suffer from any medical disease, in part of patients who had been hospitalized for a prolonged period but were neither alcoholic nor brain damaged. The experimental method corrected for individual differences in auditory acuity.

Korsakoff patients do not select randomly from environmental stimuli. If the set appropriate to the perceptual task is a single one, and is presented by explicit instruction, they can use it effectively. In general, however, they have a limited capacity for utilizing sets and for forming appropriate expectancies from cues. Their limitation is most marked when accurate perception is facilitated by maintaining two sets concurrently; they appear to be capable of carrying only one set at a time. They do, however, adopt a new set after their first expectancy is no longer confirmed, and do so with no greater difficulty than the controls, i.e., they do not persevere.

The findings were examined in relation to previous studies of attitudinal factors in Korsakoff psychosis and to Bartlett's theory of remembering.

REFERENCES

ACH, N. *Ueber die Willenstaetigkeit und das Denken*. Goettingen: Vandenhoeck & Rupprecht, 1905.

- ALLPORT, F. H. *Theories of perception and the concept of structure*. New York: Wiley, 1955.
- BARTLETT, F. C. *Remembering*. Cambridge, England: Cambridge Univer. Press, 1932.
- BUERGER-PRINZ, H., & KAILA, M. Ueber die Struktur des amnestischen Symptomenkomplexes. *Z. Neurol. Psychiat.*, 1930, **124**, 553-595.
- GRUENTHAL, E. Zur Kenntnis der Psychopathologie des Korsakowschen Symptomenkomplexes. *Mscr. Psychiat. Neurol.*, 1923, **53**, 89-132.
- LIDZ, T. The amnesic syndrome. *Arch. Neurol. Psychiat.*, 1942, **47**, 588-605.
- MOSTELLER, R., & BUSH, R. R. Selected quantitative techniques. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954. Pp. 289-334.
- PICK, A. Beitrage zur Pathologie des Denkverlaufes beim Korsakoff. *Z. Neurol. Psychiat.*, 1915, **28**, 344-383.
- SELZ, O. *Ueber die Gesetze des geordneten Denkverlaufs*. Bonn: Cohen, 1913.
- TALLAND, G. A. The effect of set on accuracy of auditory perception. *Brit. J. Psychol.*, 1958, **49**, 117-130.
- WECHSLER, D. *The measurement of adult intelligence*. (3rd Ed.) Baltimore: Williams & Wilkins, 1944.

Received December 9, 1957.

PERSONALITY AND SITUATIONAL DETERMINANTS OF CONFORMITY¹

JOHN McDAVID, JR.

Iowa Child Welfare Research Station

CONSIDERABLE evidence suggests that there are important differences among individuals in the manner in which they respond to interpersonal communications. Certain individuals tend to modify their interpretation of the communication as a function of perceived characteristics of its source, while others tend to alter their perception of the source as a function of the information he communicates. Such differences have been described in various contexts in group behavior, and their relationship to social influence phenomena has been suggested by Asch (1956) in his contrast of yielding at the *perceptual* and *action* levels, by Deutsch and Gerard (1955) in their distinction between *normative* and *informational* social influences, and by Thibaut and Strickland (1956) in their description of *group* and *task* sets in group behavior.

The present study was designed to differentiate two groups of individuals in terms of consistent dispositional tendencies in the perception of interpersonal communication. For one group, the relationship of the individual to his social group is consistently centered around the *source* of communication and, for the other, around the communicated material or *message*. These experiments then are concerned primarily with clarifying differences in the manner in which each of these groups yield or change as a function of discrepant communications from others, and secondarily with describing the psychological ingredients or personality variables that dispose an individual either to "source-oriented" or to

"message-oriented" perception of interpersonal communications. It was hypothesized (a) that the source-oriented group should yield to the judgments of others more often than the message-oriented group; (b) that yielding among the source-oriented group, who attend primarily to their relationship with other people, should be more affected by variations in the amount of discrepancy between their own position and the communicated group position and less by variations in task difficulty; and (c) that yielding among the message-oriented group, who attend primarily to the relationship between their own activity and the communicated information, should be more affected by variations in the difficulty of the task and less by variations in the amount of discrepancy between their own position and the group position.

METHOD

Subjects

Ss for these studies included 165 adolescent males between the ages of 16 and 18, all junior students in a suburban public high school. From this group, 65 were scheduled for experimental observation. The interpretational test instrument was pretested and standardized with an essentially identical sample of 120 boys from a second high school near the same city, and the design for the laboratory experiment was pretested with a sample of 100 college sophomores.

Differentiation of Source-Oriented and Message-Oriented Groups

Analysis of verbalization was employed as a technique for differentiating source-oriented and message-oriented Ss. It was assumed that an individual's free verbalization should be structured according to the perceptual or interpretational schemes he habitually employs (Whorf, 1952). A test instrument was built consisting of 10 sentence fragments depicting six particular kinds of interpersonal communication events (approval, disapproval, help, disagreement, suggestion, and doubt of one's veracity by another person) and four irrelevant events. This sentence completion test was administered with a minimum of instructions, requesting simply that *S* make a complete sentence by adding words to the fragmentary stems provided. For economy and objectivity in analyzing the verbal mate-

¹ This report is based upon a portion of a doctoral dissertation completed in 1957 in the Department of Psychology, Princeton University. The research was supported by the United States Office of Naval Research, Group Psychology Branch, under Contract NONR 171-055, H. M. Schroder, principal investigator.

The author is grateful to H. M. Schroder for criticism and suggestions, to Peter Rahm for assistance in the design and construction of electronic apparatus, to M. W. Perrine for assistance in coding sentence completions, and to the administrations of Hamilton High School, Trenton, N. J., and Pennsbury High School, Yardley, Pa., for cooperation in supplying Ss.

rial, an objective coding manual was designed and pretested on a sample of 120 Ss.² The coding manual scores completions for the six stems that refer to interpersonal events in one of two categories indicating either (a) primary concern with the message communicated (e.g., "When someone disagrees with me...I try to profit from it") or (b) primary concern with the source or communicator (e.g., "When someone disagrees with me...I try to keep them liking me"). Incomplete, dual, or otherwise ambiguous responses are classified as unscorable. Although all sentence completions for the experimental sample were coded by a single judge, a second judge coded a random sample of completions for estimation of interrater reliability, and the two judges concurred in coding over 93% of all codeable responses. In order to quantify the verbal codings, message-oriented completions were scored +1 and source-oriented completions scored -1. Since the manual applies to six stems, total scores ranged from -6 to +6. Ss obtaining a positive total score were assigned to the message-oriented group, while Ss obtaining a negative total score were assigned to the source-oriented group. The modal score was +1.

Experimental Observation of Social Behavior

In order to observe the actual behavior of individuals comprising each of the two groups (message-oriented and source-oriented), a laboratory situation involving experimental group pressure was designed. A technique previously employed to simulate group atmospheres by means of tape recordings (Olmstead & Blake, 1955) was modified for this purpose. A telephone network was arranged so that S was in two-way communication with E, with a hand telephone provided for S in a small soundproof cubicle. The entire experimental session, including instructions, materials for the behavioral task, and the simulated presence of two anonymous peers was tape recorded. Thus constancy of experimental conditions for all Ss was guaranteed. The behavioral task consisted of counting a series of sets of metronome clicks sounded at a comfortably audible volume at the rate of 200 per minute. Pretesting indicated no apparent tendency for Ss to group the clicks (i.e., to count by twos or threes) at this rate; nevertheless, all critical trials consisted of an odd number of clicks in order to minimize any intertrial differences that might thus occur. Conformity was defined here as *behavior that is patterned after a model set by other people*. Therefore, each S was informed that two of his classmates were simultaneously participating in the experiment. The instructions implied that all of S's behavior during the experiment would occur in the auditory presence of the other two Ss and E. In order to simu-

late the presence of two anonymous peers and to provide controlled models or examples for behavior, two appropriate unidentifiable voices recorded predetermined responses into the tape under circumstances identical to those in which the experimental S operated. The audio fidelity of the system was such that voices were not identifiable, and the recording could not be recognized as a tape rather than "live." Each S was informed that he was to be identified as *Subject C*, and that the other two Ss were identified as *A* and *B*. Following each trial, Ss reported in supposed alphabetical order, so that every S heard two controlled reports of the number of clicks on a trial before voicing his own report. In the instructions, S was told that his ability to use audio communication systems was being tested under military conditions where a group of men must share a single line, and that his performance was to be evaluated in terms of both his "accuracy in reporting as well as cooperation with others performing the task on this line." The nature of the task was described and an example trial was heard; subsequently two typical experimental trials were heard for unreported practice counting. Pretesting demonstrated that three practice trials were sufficient to level out differential effects of practice and experience on performance of the counting task.

The experimental sequence consisted of ten trials, of which five were controlled reinforcement trials (Tr. 1, 2, 4, 5, 8) and five were critical experimental trials (Tr. 3, 6, 7, 9, 10). In three of the five reinforcement trials (Tr. 1, 4, 8), the simulated Ss *A* and *B* were in agreement and reported the actual number of clicks; in the remaining two, *A* and *B* disagreed, but the average of their two reports was the correct number of clicks.

Two experimental task conditions were manipulated in the series, in terms of (a) difficulty of the counting task, and (b) discrepancy between the majority report and the actual number of clicks. Pretesting showed that the clicks could be counted fairly easily and accurately when the total was less than 25; however, at totals of more than 25 clicks, confidence in counting decreased rapidly though actual accuracy in counting did not drop off sharply. Therefore, two of the critical trials (Tr. 3, 7) were composed of 23 clicks and were characterized as *easy* tasks, and the remaining three were composed of 31 clicks and were characterized as *difficult* tasks. In addition, three degrees of discrepancy of majority report (model) from veridical count were employed. On two of the critical trials (Tr. 3, 9), *A* and *B* voiced identical reports that were discrepant from the actual count by one; on two trials (Tr. 6, 7) this discrepancy was two counts; and on one trial (Tr. 10) the discrepancy was three. A discrepancy of three was not presented following an easy (23 clicks) trial, since pretesting suggested that such a discrepancy was sufficiently obvious to arouse considerable suspicion among Ss as to the nature of the experiment. The entire series was run through identically and without interruption for each S. The order in which the experimental conditions occurred was randomized, and there was no evidence that any sequence or position effects influenced S's behavior on a given trial.

² The complete schedule for the sentence completion test, including the coding manual and example responses, has been deposited with the American Documentation Institute. Order Document No. 5808, from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. remitting \$1.25 for microfilm or \$ 1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

RESULTS

Yielding at Various Levels of Discrepancy and Difficulty

The procedures employed could give rise to three categories of behavior: *S* could either (a) agree with the example reports of the simulated peers, or he could (b) compromise his own count of the clicks with these example reports and report a number between the two, or he could (c) independently report his own count. Because pretesting of the experimental design showed that *Ss* could count as many as 35 clicks with over 90% accuracy, it may be assumed that an *S* who persisted in making errors on the five control trials was either uninterested or unable to perform the task adequately. For this reason, *Ss* who made three or more errors in counting the five control trials were eliminated from the sample. Of 65 *Ss* originally scheduled for observation in the laboratory, 10 failed to keep their appointments, and 7 were eliminated under the error criterion, leaving a net total of 48 *Ss* in the experimental sample. The distribution of scores derived from the sentence completion test was such that 23 of the 48 experimental *Ss* received positive scores (message-oriented group) and 21 received negative scores (source-oriented group). Four *Ss* who received zero or neutral scores were included in the source-oriented group to approximate a median split for purposes of analysis. Table 1 shows the proportion of each group of *Ss* behaving in a given manner on each of the five critical trials. It should be noted that since it is impossible to compromise with a reported discrepancy of one count, no entries occur in those cells of the table.

Two quantitative indices of conformity to group pressures were derived from the number of trials in which a given category of behavior occurred for each *S*. An Agreement Index ranging from 0 to 5 was assigned in terms of the number of trials in which *S* agreed completely with the majority report, and a Yielding Index ranging from 0 to 5 was assigned in terms of the number of trials in which *S* yielded to the majority report *either* by agreeing with it *or* by compromising with it. These data are summarized in Table 2. In order to test the significance of the difference between

TABLE 1
SUMMARY OF BEHAVIOR ON THE FIVE EXPERIMENTAL TRIALS FOR MESSAGE-ORIENTED AND SOURCE-ORIENTED GROUPS

Trial		Behavior					
Number	Description (discrepancy; difficulty)	Independent		Compromise		Agreement	
		Message-oriented	Source-oriented	Message-oriented	Source-oriented	Message-oriented	Source-oriented
3	-1; easy	39.1%	16.0%	—	—	60.9%	84.0%
6	+2; diff.	43.5	32.0	26.1%	16.0%	30.4	52.0
7	-2; easy	52.2	16.0	13.0	16.0	34.8	68.0
9	+1; diff.	69.6	68.0	—	—	30.4	32.0
10	-3; diff.	39.1	4.0	34.8	32.0	26.1	64.0

the source- and message-oriented groups on each of these indices, the Mann-Whitney *U* statistic (1947) was employed. By this test, the two groups differed on the Yielding Index at the .010 level of probability, and on the Agreement Index at well beyond the .005 level of probability.

In order to explore further the apparent differences in response of each of the two experimental groups to majority pressures, the conditional effects of task difficulty and report discrepancy upon yielding (either by agreement or by compromise) were determined for each group. Cochran's (1950) *Q* test for frequencies in related samples indicated that the effects of task difficulty are not significant for the source-oriented group, but are significant at the .02 level for the message-oriented group. Similarly, the *Q* test showed that the effects of conditional variation of discrepancy of the majority report from the actual number of clicks were not significant for the message-oriented group, but were significant at the .02 level for the source-oriented group.

In addition to this evidence of differences in the response of each group to conditional manipulations, there is some evidence that the mode of yielding behavior differs for the two groups. When yielding does occur, message-oriented individuals show some tendency to compromise rather than to agree with the discrepant majority report. A comparison of the proportion of agreement responses to total yielding responses for each group over all three trials in which both categories of behavior could occur showed that the source-oriented group agreed fully on 74.2% of all

TABLE 2

FREQUENCY OF OCCURRENCE OF TWO MODES OF YIELDING BEHAVIOR FOR MESSAGE-ORIENTED AND SOURCE-ORIENTED GROUPS

Index Score (Numbers of trials in which behavior occurred)	Agreement		Yielding (agreement or compromise)	
	Message-oriented	Source-oriented	Message-oriented	Source-oriented
0	4	1	3	0
1	8	1	2	1
2	2	7	7	1
3	6	7	4	9
4	3	6	4	9
5	0	3	3	5
<i>N</i> =	23	25	23	25

yields, as compared with 55.2% for the message-oriented group. The probability of such a difference occurring by chance in this sample is just under .10. Inspection of Table 2 reveals that of the eight *Ss* who *yielded* either by agreement or by compromise on all five experimental trials, five were included in the source-oriented group and three in the message-oriented group; however, all of the three *Ss* who *agreed* with the majority report the maximum of five times were members of the source-oriented group. These implications that members of the message-oriented group tend to compromise rather than to agree with the group are further supported by the behavior of each experimental group on Trial 10 of the series. (See Table 1.) In this trial, the majority report is maximally discrepant from the actual count (which is assumed to be the individual's own position) and the trial is characterized as difficult (31 clicks). Only one of the 25 *Ss* constituting the source-oriented group remained independent in his report, while 8 compromised with the discrepant majority report, and 16 agreed fully with it. On the other hand, 9 of the 23 *Ss* in the message-oriented group remained independent, 8 compromised with the majority, and only 6 agreed fully with it. These differences in behavior are significant at the .01 level by chi-square test.

Psychological Description of Source-Oriented and Message-Oriented Individuals.

In pursuing description of the psychological variable or variables that dispose an individual

either to message- or source-oriented interpretation of interpersonally communicated information, attention was focused upon the individual's *self concept* as an organized configuration of perceptions of himself. Although the notion of self concept is essentially that of Hayakawa (1956) and Carl Rogers the methods of this approach stem largely from the transactional (Kilpatrick, 1952) analysis of perception. Kilpatrick (1952) and Cantril have suggested that an individual's assumptions about the constancy or stability of himself are mutually related with his perception of constancy or change in his environment. It was assumed that the individual who tends to maintain a constant or stable perception of himself should be disposed to evaluate interpersonally communicated information by his own standards in terms of its utility to him, whereas the individual whose perception of himself tends to be unstable should be disposed to evaluate communicated information in terms of the stability of his perception of its source. In order to obtain data about the individual's perception of himself, a direct scaling task was devised for measurement of (a) self-evaluation and (b) self-modification. This rating task presented four hypothetical interpersonal events (help, disagreement, suggestion, and doubt of veracity) and required *S* to rate his reaction to these events along two dimensions by the method of successive intervals: (a) postevent evaluation of himself in the situation, and (b) postevent readiness to modify his perception of himself in the situation. An eight-box scalometer was provided and labeled appropriately at each end for each dimension. The self-evaluation scale was defined generally by the words "I am doing poorly" above the second box and "I am doing well" above the seventh box. The extreme ends of the scale carried brief paragraphs describing low self-evaluation at one extreme and high self-evaluation at the other. Similarly, the self-modification scale was defined generally by the words "I should change to a new approach" above the second box and "I should keep up my present approach" above the seventh box, with longer statements describing rigid persistence at one extreme and complete re-evaluation of the situation at the other. From these ratings, scale values ranging from 1 to 8 for each event were obtained, and by

summing over the four events, scores ranging from 4 to 32 were derived for each *S* on each of the two scales. The test-retest reliability measures of these scores for the pretest sample were .75 for the self-evaluation scale and .73 for the self-modification scale.

Scores on the self-evaluation and self-modification scales for the two experimental groups were compared. The mean self-evaluation score for the source-oriented group was 17.08, compared with 17.43 for the message-oriented group, a difference that is not statistically significant. The mean self-modification score for the source-oriented group was 17.32, compared with 20.52 for the message-oriented group. This difference is significant by *t* test at the .02 level ($t = 2.35$), indicating that members of the source-oriented group are more prone to modify their perception of themselves in a work situation following an interpersonal communication than are members of the message-oriented group.

DISCUSSION

A distinction between two processes underlying social conformity is supported by the findings reported here. Individuals whose interpretations of interpersonal communication events are primarily a function of the source of those communications tend to be generally more susceptible to group pressures than individuals whose interpretations are a function of the content of the communication. Furthermore, a number of specific differences in the response of two such groups clarify the description of these two processes. Members of the source-oriented group were inclined to agree with the majority position when yielding rather than to compromise with it; they yielded as often when the task was easy as when the task was difficult, but they tended to yield more when the discrepancy between the group judgment and their own judgment was relatively greater. This group, then, appears to be motivated to conform in order to avoid being different from the group. On the other hand, members of the message-oriented group showed a tendency to compromise with the majority position rather than to agree fully with it when yielding; they yielded as often to group judgments very near their own as to judgments more discrepant from their own, but they tended to yield more often as the

difficulty of the task increased. Thus, this group appears to be motivated to conform for reasons of personal utility, as a means of increasing their own chances for success. All of these findings lend credence to the hypothesis that the message-oriented individual is one who is inclined to make use of interpersonally communicated information as a tool for his own gain.

These characterizations of the two experimental groups are supported by the fact that members of the message-oriented group demonstrated less flexibility than the source-oriented group in modifying their perception of themselves in a work situation following interpersonal communication. The individual who maintains a relatively firm perceptual view of himself would very likely be inclined to shape his perceptual interpretation of communicated information in terms of its congruence with his self concept. The fact that the two experimental groups did not differ significantly in self-esteem as indexed by self-evaluation following the communication event suggests that modification or re-evaluation of oneself in a given situation is not strictly a function of one's level of self-evaluation or his ability in that situation.

SUMMARY

This study investigated the relationships between dispositional tendencies to respond to interpersonal communications in terms of either (a) the source of the communication, or (b) the message or content of the communication, and yielding to group influence as a function of (a) the amount of discrepancy between group judgment and the individual's own judgment, and (b) the difficulty of the behavioral task in which the individual is engaged. A *source-oriented* group and a *message-oriented* group of *Ss* were differentiated on the basis of their habitual manner of interpreting interpersonally communicated information by means of objective analysis of sentence completions. *Ss* were observed in an experimental pseudo-social situation in which two anonymous peers reported judgments of an auditory counting task before the experimental *S* reported his response. The effects of task difficulty and report discrepancy were compared for each group. The message-oriented group differed from the source-oriented group in that

members of this group (a) were generally less susceptible to group influence, (b) were less affected by manipulations of report discrepancy and more affected by manipulations of task difficulty, and (c) showed a tendency to compromise with discrepant group judgments rather than to agree completely with them when yielding did occur. In rating their own reactions to interpersonal communication events, the source-oriented individuals indicated significantly greater flexibility in modifying their perception of themselves following that event than the message-oriented individuals, although the two groups did not differ in level of postevent self-evaluation.

REFERENCES

- ASCH, S. E. Studies of independence and submission to group pressure. I. A minority of one against a unanimous majority. *Psychol. Monogr.*, 1956, **70**, 9 (Whole No. 416).
- COCHRAN, W. G. The comparison of percentages in matched samples. *Biometrika*, 1950, **37**, 256-266.
- DEUTSCH, M., & GERARD, H. B. A study of normative and informational social influences upon individual judgment. *J. abnorm. soc. Psychol.*, 1955, **51**, 629-636.
- HAYAKAWA, S. I. The semantic barrier: A theory of personality and communication. Providence, R. I.: Walter V. Clarke Associates, 1956. (Pamphlet)
- KELMAN, H. C. Three processes of acceptance of social influence: Compliance, identification, and internalization. *Amer. Psychologist*, 1956, **11**, 361. (Abstract)
- KILPATRICK, F. P. (Ed.) *Human behavior from the transactional point of view*. Hanover, N.H.: Inst. for Associated Research, 1952.
- MANN, H. B., & WHITNEY, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50-60.
- McDAVID, JOHN, JR. Conceptual processes in social conformity. Unpublished doctoral dissertation, Princeton University, 1957.
- OLMSTEAD, J. A., & BLAKE, R. R. The use of simulated groups to produce modifications in judgment. *J. Pers.*, 1955, **23**, 335-345.
- THIBAUT, J. W., & STRICKLAND, L. H. Psychological set and social conformity. *J. Pers.*, 1956, **25**, 115-129.
- WHORF, B. L. *Collected papers on metalinguistics: The relation of habitual thought and behavior to language*. Washington, D. C.: Department of State, Foreign Service Institute, 1952.

Received December 12, 1957.

THE RELATIONSHIP BETWEEN COPING AND AVOIDING BEHAVIOR AND RESPONSE TO FEAR-AROUSING PROPAGANDA

MICHAEL J. GOLDSTEIN^{1, 2}

University of Washington

IN THEIR comprehensive study of the effectiveness of fear-arousing propaganda, Janis and Feshbach (1955) attempt to explain the failure of the strong fear appeal by recourse to the concept of a "defensive reaction." They suggest that the arousal of excessive amounts of emotional tension may stimulate in subjects a "defensive reaction" that is incompatible with the recall, at a later date, of the recommendations contained in the propaganda. However, recent research on the defense mechanisms (Carlson, 1954; Carpenter, Weiner, & Carpenter, 1956; Eriksen: 1952a, 1952b, 1954; Mainord, 1956) suggests that there is no uniform defensive reaction to the heightening of emotional tension. In the perceptual area, for example, some subjects (Ss) may react with a lowered threshold for tension-producing material, and others with a heightened threshold (Carpenter et al., 1956; Eriksen, 1952a). With memory tasks, some Ss tend to recall tension-producing material better than neutral material, while for others the reverse is true (Carlson, 1954; Eriksen, 1952b; Mainord, 1956). The defensive reaction of *S* seems to reflect a habitual way of dealing with tension-producing material.

The present study applies this view of individual differences in defensive reactions to the type of response shown to fear-arousing propaganda. It is hypothesized that the acceptance or nonacceptance of the recommendations contained in a propaganda appeal is related to the *S*'s characteristic reaction to tension-producing stimuli. If an *S* tends to recall neutral stimuli better than tension-producing stimuli, he should respond more favorably to a minimal fear appeal, in which

the recommendations are introduced in a relatively neutral context. If, on the other hand, he tends to recall tension-producing stimuli better than neutral stimuli, he should respond more favorably to a strong fear appeal, in which the recommendations are introduced in a setting of heightened emotional tension. It is hypothesized that certain "defensive reactions," in the sense of Janis and Feshbach (1955) facilitate the acceptance of propaganda, while others interfere with its acceptance. Response to propaganda is thus construed as an interaction between the person's characteristic mode of responding to tension-arousal and the level of tension stimulated by the appeal.

In order to test this hypothesis, it is necessary to select Ss, prior to their exposure to the propaganda, on the basis of their characteristic mode of responding to tension-arousing stimuli. Mainord (1956) has recently introduced a relevant distinction between copers and avoiders, in terms of Ss' responses to a highly emotionally-charged form of the Sentence Completion Test (SCT). Persons who are sensitized to the sexual and aggressive implications of the various sentence stems and can relate them to their own needs and emotions are identified as copers; avoiders, on the other hand, are those who characteristically fail to recognize the implications of the sentence stems and are unable to perceive the feelings suggested in the stems as being related to themselves. The distinction parallels that drawn in perceptual situations between persons who show sensitization or vigilance and those who engage in perceptual defense. The two groups appear to utilize different techniques for defending against the arousal of anxiety. Mainord's experiment suggests that these categories permit the effective prediction of the recall of disturbing and neutral stimuli. Using a paired-associate task in which nonsense syllables were paired with neutral or with disturbing words, Mainord found that copers recalled more nonsense syllables that had been

¹ A portion of a dissertation submitted to the Department of Psychology, University of Washington (Goldstein, 1957). The author would like to express his gratitude to Allen Edwards for his support and guidance throughout the course of this study. He would also like to express his thanks to Hayden Mees, William O'Donnell, Vitale Rozyneko, James Taylor, and Gary Clune for their help.

² Now at the University of California at Los Angeles.

associated with disturbing words than ones associated with neutral words, while in the case of avoiders, the advantage lay with syllables associated with neutral words. There was almost no overlap between the two groups in their recall behavior, in spite of the fact that the groups did not differ on the initial learning task.

The present study compares the response of copers and avoiders to fear-arousing propaganda. Since Mainord's results suggest that copers excel avoiders in recall for disturbing stimuli, it is hypothesized that acceptance of the recommendations contained within a strong fear appeal is greater among copers than among avoiders. The evidence for superior recall for neutral material among avoiders suggests that a minimal fear appeal, in which the recommendations are presented in a relatively neutral context, should gain greater acceptance among avoiders than copers. These hypotheses predict an interaction between personality type and level of fear arousal.

METHOD

The present experiment employed as Ss the entire freshman class at a medium-sized high school in the State of Washington.³ All materials were presented in Health Education classes and were conducted as part of the regular Health Education curriculum.

Selection of Copers and Avoiders

The SCT was administered to the total pool of 350 Ss on the first experimental day. The test consists of 60 sentence stems, of which 40 are critical items and 20 are filler items. The items are so arranged that each group of two critical items is followed by one filler item. Mainord had selected copers and avoiders on the basis of scores assigned to the responses to the critical stems, each of which was scored 0, 1, or 2 on the following criteria:

1. The more specific the response the higher the score
2. The stronger the expressed feeling, the higher the score
3. The more arbitrary the response, the lower the score

A person's score on the coping-avoiding dimension is merely the sum of the weights assigned to the individual items. Applying this scoring system to a random sample of 50 records, the present experimenter and another clinical psychologist obtained a correlation of .62 between their ratings, indicative of unsatisfactory

reliability. In an effort to increase reliability by providing a more operational criterion of the coping-avoiding dimension, a manual was constructed giving sample responses for each scoring weight for all critical sentence stems. Sample critical stems are presented below with criterion examples for each of the three scoring categories:

1. If I were struck:
 - [2] I would hit back
I would get mad
 - [1] I'd quit
I would call for help
 - [0] By lightning, I would die
I don't know
2. The worst thing a girl can do:
 - [2] Sell herself or go willingly
Think about a male's sex
Have a baby before she is married
 - [1] Lie
Slap a boy
Be stuck up
 - [0] Go to a beauty parlor
Eat too much
Not be ladylike
3. I hate:
 - [2] My parents
Mr. Jones
My sister
 - [1] Some people
Democrats
Being called names
 - [0] Snakes and wiggly worms
Pickles
Nothing
4. A girl's figure:
 - [2] Is very important to me
Is to have fun with
Hard to keep your eyes off
 - [1] Has a lot to do with friends
Should be feminine
Are pretty good
 - [0] Is slim
Is not their personality
I don't know

This manual was presented to graduate students in clinical psychology who used it in scoring subsamples of 50 records each. The correlations between ratings by the investigator and those of the other raters were: .92, .90, .90, .92, .88, and .89, indicating that the records had been scored in a reliable fashion. Mean scores based on two raters' scoring were therefore used in selecting Ss for the experiment. A comparison of scores for males and females indicated that there were significant differences in both mean and variance between the two groups (males $\bar{x} = 32.75$, $SD = 7.30$, females, $\bar{x} = 34.68$, $SD = 9.25$). For this reason copers and avoiders were selected separately from the distribution of scores for each sex. Copers were selected from the upper 25% of the distribution, while avoiders were selected from the bottom 25%, yielding 38 males and 40 females in each extreme group. Absences at one of the later experimental sessions reduced the total to 67 copers and 72 avoiders.

³ The author is indebted to Don Hartung, Principal and the Health Education Staff of Puyallup High School, Puyallup, Washington for their help in providing the Ss for the present study.

Propaganda Materials and Method

The propaganda materials in the present study were identical to those used in a previous study by Moltz and Thistlethwaite⁴ (1955), and represent a slight modification of the original appeals dealing with dental hygiene which were developed by Janis and Feshbach (1955). The two talks, a strong fear appeal and a minimal fear appeal, differed in the number of threat references and in the threatening nature of the accompanying slides. Both appeals were recorded on tape by a local radio announcer and were regarded as thoroughly professional by all who heard them. The tapes and slides were presented to the Ss in their regular hygiene classes. Five classes received the strong fear appeal and six received the minimal fear appeal, following a simple alternation procedure. All Ss received the talks during the course of a single day, which occurred approximately one week after the administration of the SCT.

In an effort to assess the effectiveness of the propaganda, three questionnaires were used: (a) a pretest questionnaire administered one week prior to the appeals, (b) a posttest questionnaire, administered immediately after the appeal was concluded, and (c) a retention questionnaire, administered two weeks after the appeals were presented. These questionnaires contained the following various subscales:

Dental practices scale. Five items dealing with current tooth-brushing practices. These practices were the ones which the propaganda was designed to modify. This scale appeared in the pretest and two-week retention questionnaires. A preliminary analysis of scores on this scale in the pretest questionnaire indicated that they were uniformly low and did not differ significantly for copers and avoiders. It was possible for each S to show either an increase, decrease, or no change in the number of correct practices after exposure to propaganda.

Dental information scale. Twenty-two items dealing with both the factual information concerning causes of tooth decay and various recommendations contained in the lecture. Each S was given a score on the basis of his number of correct answers. The scale appeared in all three questionnaires.

Dental anxiety scale. Sixteen items dealing with anxiety related to diseases of the mouth and gum. On five of the items, Ss were asked to rate on a five-point scale the degree of anxiety that he was experiencing about dental matters; on the remaining items, the rating was on a three-point scale. In scoring, each item was reduced to a two-point scale by dichotomizing the possible responses to each item as close as possible to the median. Elimination of one item which Ss had difficulty in handling resulted in a scale with a potential range from 0 to 15. The scale appeared in both the pretest questionnaire and the posttest questionnaire.

Each of the three questionnaires was embedded in an entirely different set of filler items designed to mask the primary interest of the study.

⁴ The author would like to express his thanks to I. L. Janis, H. Moltz, and D. Thistlethwaite for providing copies of the propaganda materials.

RESULTS

The effectiveness of the propaganda appeals was evaluated on the basis of changes from pretest to two-week retention test in Ss' responses to the five-item dental practices scale. A S could show one of three possible responses: 1. An increase in the number of reported correct practices (*positive conformity*). 2. No shift in the number of correct practices (*no change*). 3. A decrease in the number of correct practices (*negative conformity*). In order to take into account alienation by the propaganda as well as favorable reaction to it, the measure of effectiveness used in the present study was the percentage of Ss showing positive conformity minus the percentage showing negative conformity. A discussion of this measure of *net effect*, together with a description of tests of significance for it, are contained in Hovland, Lumsdaine, and Sheffield (1949, p. 304).

Net Effect of Strong Appeal

An analysis of the net effect for each sex revealed no significant differences between males and females. Therefore, further analyses employed the combined scores for both sexes in the copers and avoider groups. The net effect of the strong fear appeal on copers and avoiders is presented in Table 1. It was predicted earlier that the strong fear appeal would be more effective with copers than with avoiders. The obtained net effects of 26% for copers and 8% for avoiders are in the predicted direction; however, the critical ratio for the difference in net effect is 1.00, which is not significant at the .05 point.

Net Effect of the Minimal Appeal

An opposite prediction was made for the minimal fear appeal, namely, that this appeal should be more effective with avoiders than

TABLE 1
NET EFFECT OF STRONG FEAR APPEAL ON COPERS AND AVOIDERS

Response	Copers		Avoiders	
	%	N	%	N
Positive conformity	43	15	34	14
No change	40	14	39	16
Negative conformity	17	6	26	11
Net Effect	+26		+8	

TABLE 2

NET EFFECT OF MINIMAL FEAR APPEAL ON COPERS AND AVOIDERS

Response	Copers		Avoiders	
	%	N	%	N
Positive conformity	44	14	55	17
No change	40	13	35	11
Negative conformity	16	5	10	3
Net Effect	+28		+45	

with copers. Table 2 presents the responses of the two groups to the minimal appeal. The net effect of the minimal fear appeal is also in the predicted direction, but once again the critical ratio for the difference in net effect ($C.R. = 1.00$) does not approach statistical significance.

Interaction Between Personality Type and Level of Fear Arousal

The critical test of the present hypothesis, however, involves the second order comparison in which the reactions of copers and avoiders to both types of propaganda appeals are considered together. This second order effect is computed as follows:

	Copers	Avoiders	Difference
Strong fear appeal	26%	8%	= 18%
Minimal fear appeal	28%	45%	= -17%
Difference between the differences			35%

The critical ratio of this second order effect is 1.60, which is significant at the .05 point for a one-tailed test of significance.⁵ This finding of a significant interaction between the method of dealing with tension-producing stimuli and the level of tension arousal intended by a propaganda appeal supports the basic hypothesis of this study.

It is interesting to note that the significant interaction is due almost exclusively to the marked difference in the response of the avoider group to the two lectures. Both lectures were relatively ineffective with the copers, a fact that does not confirm the hypothesis that a strong fear appeal would be maximally effective with this group. This

⁵ Special thanks are due Fred Sheffield for informing the author of the method of extending the test of significance for net effect to the second order comparisons.

result parallels the finding of Feshbach and Janis (1954) who noted that Ss high in anxiety showed a marked discrepancy in their response to a strong and a weak fear appeal, while neither appeal was particularly effective with those low in anxiety. The parallel findings raise a question as to possible equivalence of the distinction between copers and avoiders and that between Ss low and high in anxiety. A copy of the original anxiety scale used by Feshbach and Janis was contained in the pretest questionnaire in the present study, and it was therefore possible to examine the anxiety scores of copers and avoiders in each lecture group. Although there appeared to be some differences in the anxiety scores between the groups that received the strong and weak appeals, these differences were not significant at the .05 level of confidence. Also, the differences between the means of the copers and avoiders (copers $\bar{x} = 23.50$, $SD = 6.26$, avoiders $\bar{x} = 22.84$, $SD = 5.84$) when viewed independently of the lecture to which they were assigned were not statistically significant. Thus, while the patterns of results in both the present study and the Feshbach and Janis study are highly similar, they are not explicable by recourse to a common factor.

Recall of Lecture Content

Janis and Feshbach (1955) had suggested that the failure of the strong fear appeal to receive acceptance was due to the inability of Ss to recall the recommendations at a later date when appropriate action was possible. In an effort to test this hypothesis, the score of copers and avoiders on the dental information scale were investigated. If recall of content is a relevant factor in determining acceptance of the recommendations contained in an appeal, then the Ss who accept the recommendations should also recall more of the content than Ss who do not accept them. An analysis of variance of the dental information scores failed to demonstrate any significant F s either for the learning condition or for the scores obtained on the two-week retention test. Thus, the hypothesis that differences in acceptance behavior are mediated by differential recall of the content of the appeals is not substantiated.

Anxiety Stimulated by Appeals

The question may be raised of whether the copers and avoiders show a differential sensitivity to tension-producing stimuli. Do copers, for example, respond with less fear to a strong fear appeal than avoiders, and can such a difference be used to explain the differential conformity of the two groups to the two appeals? The dental anxiety scale, it will be recalled, was administered twice, one week prior to the administration of the appeals and once again immediately after the appeals. Since the pretest analysis revealed a significant difference between the male and female copers who later received the strong fear appeal, an analysis of covariance of the factorial design was undertaken in the manner described by Lindquist (1953). The pretest dental anxiety scores were used as the control variable to adjust the postlecture sums of squares. This analysis indicates that copers and avoiders did not react with different degrees of anxiety to the two appeals. It cannot be concluded, on the basis of this analysis, that differences in tension threshold can explain the differential acceptance patterns of copers and avoiders.

DISCUSSION

It should be emphasized that the index of propaganda acceptance in the present study is verbal conformity. Although it is of critical importance that verbal conformity be related eventually to action taken outside of the experimental situation, the present study deals merely with changes in verbal behavior as indexed by responses to questionnaires. Within this limited area of study, the obtained results tend to substantiate the major hypothesis. The two levels of fear-arousal do stimulate differential acceptance among copers and avoiders.

The superiority of the minimal fear appeal with avoiders is clearly demonstrated; however, it is readily apparent that the copers do not respond particularly well to either appeal. Since copers show the capacity, on the basis of their SCT responses, to respond readily to a marked degree of emotional stimulation, it is possible that the strong fear appeal did not constitute an adequate level of emotional arousal for that group. Perhaps copers require

a severe threat to their sense of well-being before reacting in the predicted manner.

It is only in the avoider group that a consistency in response pattern is observed. An inability to respond directly to the tension-producing stimuli of the SCT is paralleled by an inability to respond favorably to a strong fear appeal. If avoiders can be thought of as using a habitual "defensive reaction" to tension-producing stimuli in the sense of Janis and Feshbach (1955), then this "defensive reaction" has been demonstrated in response to two sets of stimuli of widely differing content. The consistency in response pattern suggests that the concept has a certain degree of validity in explaining the basis for the failure of the strong fear appeal. Its limitation lies in the fact that it applies only to *Ss* whose habitual "defensive reaction" is of the avoidance type. For copers the explanation does not hold. Rather than viewing "defensive reactions" as a single class of behavior, it seems preferable to distinguish between different *types* of "defensive reactions," some of them facilitating and others interfering with the acceptance of fear-arousing propaganda.

Although the prediction of differential acceptance behavior by copers and avoiders was based on assumption of differential recall patterns, the analyses of learning and retention of content provide no evidence of differential recall. The mediating mechanism for the responses of copers and avoiders thus remains obscure. This finding is particularly interesting in view of the weight originally placed by Janis and Feshbach (1955) upon the role of recall in explaining the failure of the strong fear appeal. If there is no relationship between the recall of the content of a propaganda appeal and its acceptance, then it becomes necessary to search for other explanations for the basis of acceptance behavior.

Although an interaction between a particular personality variable and the level of emotional tension stimulated has been found in the present study, it leaves the conclusions of Janis and Feshbach largely unchanged. A minimal fear appeal is still most effective in eliciting acceptance of propaganda. It is not particularly effective with persons classed as copers yet it does not alienate, as does the strong fear appeal, the avoiders. There does not appear to be

any great advantage in tailoring the level of fear stimulation contained in a propaganda appeal to different personality types if it turns out that a low level of emotional tension is most effective for all personality types. Unless some form of propaganda can be found that is particularly effective with the copers groups, a minimal fear appeal still stands as the best bet for the propagandist.

SUMMARY

The present study was concerned with the response to fear-arousing propaganda of two classes of persons who characteristically use different means of dealing with tension-producing material. The two classes, "copers" and "avoiders," were selected on the basis of responses to the SCT. Avoiders are unable to recognize tension-producing stimuli and relate them to themselves, while copers show the capacity to recognize such stimuli and relate them to themselves. Each *S* was exposed to one of two propaganda lectures, a strong fear appeal or a minimal fear appeal. These lectures differed in number of threat references and in the nature of the accompanying slides.

The results support the hypothesis that a strong fear appeal receives greater acceptance among copers than among avoiders, while the minimal fear appeal receives greater acceptance among avoiders than among copers. The obtained pattern of acceptance is largely due to the marked differential effectiveness of the two appeals on the avoider group, and not as was originally predicted, to any differential acceptance by the copers.

Differences in the effectiveness of the two appeals could not be attributed to differential recall of the content of the lectures. The mediating mechanism for the difference in effectiveness thus remains unclear.

It is suggested that the concept of a "defensive reaction" to the arousal of strong

emotional tension, introduced by Janis and Feshbach, is limited in its explanatory power. The present study indicates that it may be valuable to discriminate between different types of "defensive reactions" in attempting to predict and explain the eventual acceptance or rejection of a propaganda appeal.

REFERENCES

- CARLSON, V. R. Individual differences in the recall of word association test words. *J. Pers.*, 1954, **23**, 77-87.
- CARPENTER, B., WEINER, M., & CARPENTER, JANETH. Predictability of perceptual defense behavior. *J. abnorm. soc. Psychol.*, 1956, **52**, 380-383.
- ERIKSEN, C. W. Defense against ego threat in memory and perception. *J. abnorm. soc. Psychol.*, 1952a, **47**, 230-235.
- ERIKSEN, C. W. Individual differences in defensive forgetting. *J. exp. Psychol.*, 1952b, **44**, 442-446.
- ERIKSEN, C. W. Psychological defenses and "ego-strength" in the recall of completed and uncompleted tasks. *J. abnorm. soc. Psychol.*, 1954, **49**, 45-50.
- FESHBACH, S., & JANIS, I. L. Personality differences associated with responsiveness to fear arousing communications. *J. Pers.*, 1954, **23**, 154-166.
- GOLDSTEIN, M. The relationship between coping and avoiding behavior and attitude modification. Unpublished doctoral dissertation, Univer. of Washington, 1957.
- HOVLAND, C. I., LUMSDAINE, A. A., & SHEFFIELD, F. D. *Experiments in mass communication*. Princeton, N. J.: Princeton Univer. Press, 1949.
- JANIS, I. L., & FESHBACH, S. Effects of fear arousing communications. *J. abnorm. soc. Psychol.*, 1955, **51**, 663-668.
- LINDQUIST, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
- MAINORD, W. A. Experimental repression related to coping and avoidance behavior in the recall and relearning of nonsense syllables. Unpublished doctoral dissertation, Univer. of Washington, 1956.
- MOLTZ, H., & THISTLETHWAITE, D. Attitude modification and anxiety reduction. *J. abnorm. soc. Psychol.*, 1955, **50**, 231-237.

Received December 17, 1957.

A COMPARISON OF DIRECT, INDIRECT, AND FANTASY MEASURES OF IDENTIFICATION¹

JAMES BIERI, ROBIN LOBECK

New York School of Social Work, Columbia University

AND M. DAVID GALINSKY

University of Michigan

MUCH discussion has centered about the theoretical formulation of identification in terms of its relationship to introjection (Knight, 1940), aggression (Freud, 1946), and learning (Seward, 1954). A recurrent controversy is whether identification is best construed as a normal process of becoming like a significant other through social learning, or whether identification should be considered defensive behavior stemming from an interpersonal conflict (Sanford, 1955). Our own view is that it is worthwhile to keep in mind two necessary (but not sufficient) aspects of identification. First, some degree of *similarity* between the behavior of a subject and the object of identification is necessary. This aspect of identification is emphasized by those who view identification as a developmental process of social learning (Mowrer, 1953; Seward, 1954), and by those who measure identification empirically in terms of the real or perceived similarity between the object and the subject (Beier & Ratzeburg, 1953; Cava & Raush, 1952; Lazowick, 1955; Livingstone, 1956; Payne & Mussen, 1956; Sopchak, 1952). A second aspect of identification is the condition that some *degree of involvement* or closeness of relationship must exist or have existed between the object and the subject. Those who emphasize the defensive nature of identification tend to stress this aspect of the concept (Freud, 1946; Sanford, 1955).

Each of these aspects of identification may be assessed with varying degrees of *directness*. By degree of directness we mean the extent to which a person is aware of and focused upon

reporting an aspect of his experience, in this case, his relationships with his parents. Thus, we used the interview as a direct technique with which to measure both perceived similarity and involvement as aspects of identification. As an indirect measure of perceived similarity we used Osgood's semantic differential (Osgood, Suci, & Tannenbaum, 1957). Finally, we used the Thematic Apperception Test (TAT) as a fantasy measure (very indirect) of involvement.

Many research studies of identification in recent years have used *indirect* measures of identification. Most of these indirect measures assess identification by the degree of similarity between the way a subject (S) fills out an inventory for himself and the way he fills it out for one or both of his parents (Beier et al., 1953; Cava et al., 1952; Livingstone, 1956; Sopchak, 1952). Other studies using this indirect approach have compared S's responses with those of his parents (Lazowick, 1955; Payne & Mussen, 1956).

Our purposes in this research are twofold. First, we wish to extend the range of empirical measures of identification to include direct and fantasy measures, as well as the more common indirect measures, and to analyze systematically the consistency among these measures. Second, we wish to include several measures of each aspect of identification (i.e., similarity and degree of involvement) and to examine the relationship between these measures. These considerations lead to the following general predictions:

Hypothesis 1. There will be significant positive relationships among the various empirical measures of the perceived similarity aspect of identification.

Hypothesis 2. There will be significant positive relationships among the empirical measures of the involvement aspect of identification.

Hypothesis 3. There will be significant posi-

¹ This study was supported by Research Grants M-1274 and M-1976 from the National Institute of Mental Health, of the National Institutes of Health, U. S. Public Health Service. Facilities for the research were provided by the Laboratory of Social Relations, Harvard University, and the Research Center, New York School of Social Work, Columbia University. The writers wish to acknowledge the assistance of Wendy M. Bradburn in the collection of data.

tive relationships between the empirical measures of the perceived similarity and involvement aspects of parental identification.

METHOD

Subjects

Results for two groups of *Ss* are presented. The Female group consisted of 30 college undergraduates attending a summer school session of Harvard University. There were two married women in this group. The age range was from 18 to 24 years, with a median age of 19.4 years. The Male group consisted of 60 college students, including 30 attending Harvard College and 30 summer school students at the same institution. Among the men, there were three married *Ss*, while the group had an age range of 17 to 29 years, with a median age of 20.5 years. All *Ss* were paid at existing research rates.

As a result of the intensive nature of the assessment, each *S* was seen in two experimental sessions. The semantic differential was administered during the first session, while the TAT and interview were reserved for the second session.

Empirical Measures of Identification

Direct measure. A structured interview centering about *S's* relationships with his parents formed the basis of the direct measurement of both aspects of identification. This interview consisted of a series of 12 questions discussed with *S* in a prearranged order. Content for these interview items were suggested in part by the interview schedule used by Adorno, Frenkel-Brunswik, Levinson, & Sanford (1950). All interviews were tape-recorded, with *S's* consent, and were subsequently transcribed verbatim for scoring purposes.

Two interview questions formed the basis of the direct measures of identification:

1. Which parent do you feel you are most like at the present?

2. Which parent did you feel closer to as a child?

The first item ("like") was the direct measure of perceived similarity as an aspect of identification. The second item ("child") was the direct measure of the degree of involvement. The interviewer attempted to obtain as much amplification of each response as was deemed necessary for our scoring purposes. Each of these two questions was interspersed among other items that were not used in this study but had relevance to the parent-child relationship, including topics such as discipline, satisfactions, and conflicts. The interviews varied from 15 to 30 minutes in length, and were all ways conducted at the end of the last research session so that the effects of discussing this direct material on the indirect and fantasy measures would be minimized.

The over-all responses to the two interview items were scored in terms of one of three general categories: (a) the response clearly indicated one parent as being most relevant to the query (scored M or F); (b) both parents were mentioned as relevant to the question, but one was given greater weight than the other (scored M/F or F/M); (c) neither parent was seen as crucial or relevant to the question (scored 0), or *S* could not

make a choice in the direction of either parent (scored ?).

Responses in Category (a) are exemplified by a statement such as "I'm definitely most like my father" ("like" item) or, "I was closer to my mother as a kid" ("child" item). Category (b) was scored for statements such as, "I'm really like both my parents in some ways, but I guess I'm somewhat more like my mother" ("like" item). Responses in Category (c) might include statements as "I'm like both, but I don't think I'm like one more than the other" ("like" item) or, "I wasn't particularly close to either of my parents as a youngster" ("child" item).

Considering men and women together, 67% of the responses to both interview items fell in Category (a), 20% fell in Category (b), and 13% were scored in Category (c). It is evident that these three categories represent decreasing degrees of decisiveness with which *S* verbalizes the relative importance of one parent compared to the other in relationship to himself. The fact that 87% of all *Ss* gave responses in Categories (a) or (b) indicates that such a direct measurement of identification is feasible, whatever its subsequent validity may prove to be.

All scoring of the interviews was done by one person. Interrater reliability was assessed by having a second person independently score 20 randomly selected interviews. Of the 40 responses scored, 87% agreement was found. All five disagreements in scoring were caused by a confusion of Categories (a) and (b). Thus, when these two categories were combined (e.g., M/F responses were considered to be M responses), 100% agreement was achieved.

Indirect measure. The indirect measure of the perceived similarity aspect of identification was Osgood's semantic differential (Osgood et al., 1957). Essentially, this is a seven-point rating scale, each end of the scale being one of a pair of bipolar adjectives. *S* is asked to rate a person on all nine pairs of adjectives. Adequate reliability of ratings has been demonstrated (Osgood et al., 1957). Our list of adjectives is the same as that used by Lazowick (1955), and contains nine adjective pairings derived from Osgood's factor analysis so as to contain three pairings on each of the three most general factors obtained. Thus, three pairings have to do with potency (strong-weak, heavy-light, and rugged-delicate), three with evaluation (clean-dirty, fresh-stale, and happy-sad), and three with activity (fast-slow, hot-cold, and active-passive). *S* was asked to rate himself, his father, and his mother on each of these nine dimensions. There was a random ordering of the dimensions for each person rated so as to reduce possible response sets in the ratings. A score of perceived similarity to each parent was obtained by summing the squares of the differences between the rating *S* gave himself and that he gave a parent on every adjective pairing. For example, a high score for the mother (M score) reflects a perception of greater dissimilarity between *S* and the mother, while a low M score reflects greater perceived similarity between *S* and the mother. A second identification score was obtained from the semantic differential by subtracting the M score from the F score, yielding a difference score (D score). *Ss* with positive D scores were considered to perceive themselves as more like the mother, and *Ss* with

negative D scores were considered to perceive themselves as more like the father.

We consider the semantic differential to be an indirect measure of identification because S is not making a direct and deliberate comparison of himself with one parent in contrast to the other as in the interview.

Fantasy measure. The use of the TAT represented our most extreme indirect assessment of identification. Here we were concerned with the degree of involvement S manifested in his relationships with each parent. Our reasoning in arriving at an objective scoring scheme for the TAT measure of identification involved one central assumption: that in the stories an individual tells to cards depicting an older and younger person in close proximity, the stronger the influence of the parent-child relationship in his current behavior, the more likely that his stories depict an explicit parent-child relationship. Thus, we selected two cards that could be construed as involving a parent-son relationship for males, and two that could depict a parent-daughter relationship for females.

For male Ss, we used from the standard TAT set Card 7BM (older man and younger man) and Card 6BM (older woman and man, commonly called the mother-son card). Standard TAT instructions were given on these two cards, designed to tap involvement in the father-son and mother-son relationships respectively. For female Ss, only Card 7GF of the standard TAT cards was deemed suitable for investigating this aspect of parental identification. This card depicts an older woman sitting on a sofa speaking or reading to a girl next to her. To obtain an appropriate picture for the father-daughter relationship, a modification of the Make-a-Picture-Story Test (MAPS) was devised. A card from the MAPS was selected depicting an empty room containing a closet. To one side was placed a female figure of adolescent or young adult age, while on the other side of the room a figure of an older man was placed. As in the case of males, standard TAT instructions were given for both cards.

The TAT stories were tape-recorded and scored from the transcriptions. Each story was scored on a six-point scale so arranged as to reflect the degree to which the story explicitly contained a parent-child relationship. The following criteria were used in scoring:

- 5: parent-child relationship explicitly stated in which there is close personal involvement with each other;
- 4: parent-child relationship with a minimum of personal involvement;
- 3: two persons who are relatives (but not parent and child) or close personal acquaintances;
- 2: two persons not related but living in close physical proximity (e.g., landlady and tenant), or two persons with a business and professional relationship (e.g., lawyer and client);
- 1: two persons together in the story only incidentally (e.g., man delivering message to woman);
- 0: only one person mentioned in the story, or both figures are the same person (e.g., young man thinking when he is older).

Such a scoring scheme obviously ignores many of the subtle cues contained in thematic stories. However, it does systematically order the stories in terms of a gradient of explicit parent-child involvement. It has the further advantage of making scoring relatively ob-

jective, as reflected in the analysis of interrater reliability. Of 60 stories independently scored by two persons (one of whom scored the stories for all Ss), there was perfect agreement on 48 (80%) of the stories. Further, of the 12 stories on which disagreement existed, 11 scores deviated by a value of only one.

These scores were used in two ways as measures of the involvement aspect of identification. First, all Ss' scores on a particular card could be compared to indicate degree of involvement. Secondly, scores between two cards could be compared so as to indicate the relatively greater involvement with one parent or the other. For this latter purpose, we simply subtracted the score S obtained on the Father card from the score obtained on the Mother card (M-F). We then assumed that the higher the (M-F) score, the greater is the involvement with the mother as compared to the father.

RESULTS

On each of the identification measures used, men and women had very comparable ranges of scores. Since scores on some tests were continuous in nature (semantic differential M and F scores, and TAT) while others were discontinuous (semantic differential D scores and interview), the relationships stated in our general hypotheses were tested by chi square. The number of Ss used in comparing one measure to another vary in the ensuing tables because none of the neutral scores (those reflecting no parental preferences in response) was used. On continuous measures, scores were dichotomized as close to the median as possible. The interview was dichotomized by grouping responses in the first two scoring categories ([a] and [b] above) on each parent. By combining these two categories we were not only able to maximize the number of Ss used in our analyses, but in addition utilized the 100% scoring reliability for these grouped categories, reported above. One exception was made to this in the case of the question, "which parent did you feel closer to as a child?" Because only nine male Ss were scored for "father," we included in our analysis the nine additional cases in which replies were scored in Category (c), i.e., neither parent.

Hypothesis 1

Our first analysis concerned the prediction that measures of the perceived similarity aspect of identification will be positively associated with each other. Table 1 presents findings comparing the interview "like" question ("Which parent do you feel you are most like at present?") with the three scores of the

TABLE 1
COMPARISON OF PERCEIVED SIMILARITY MEASURES OF
IDENTIFICATION
(Chi-square analysis)

Interview "like" Item	Semantic Differential					
	F score		M score		D score	
	High	Low	High	Low	Plus	Minus
Female Group						
Mother	8	8	6	10	9	7
Father	5	6	9	2	1	10
<i>p</i>	—		.03		.02	
Male group						
Mother	14	8	10	12	15	7
Father	12	21	13	20	12	21
<i>p</i>	.05		—		.02	

semantic differential.² The results for the Female and Male groups indicate that two of the three measures of identification on the semantic differential are significantly associated in the predicted direction with the "like" item on the interview.

Hypothesis 2

The second analysis centered about the prediction that measures of the involvement aspect of identification will be positively associated with each other. Table 2 compares the findings for the "child" item on the interview ("Which parent did you feel closer to as a child?") with the TAT measures of identification. While none of the predicted relationships was significant for the Female group, it is to be noted that for males two TAT measures of the involvement aspect of identification were significant when compared with responses to the "child" item of the interview. It should be recalled that these findings for men compare those men who report they were closer to the mother as a child with those who report they were closer to the father or to neither parent. When the "neither" category is omitted, no significant relationships are found for the Male group. Thus, our findings for men must be interpreted in terms of males who either were or were not closer to the mother as a child.

² All probability values in the tables are for a one-tailed test, as the direction of the relationships was predicted. For females, Fisher's exact probability test was used. For males, chi square corrected for continuity was used.

TABLE 2
COMPARISON OF INVOLVEMENT MEASURES OF
IDENTIFICATION
(Chi-square analysis)

Interview "child" Item	TAT					
	Mother Card		Father Card		(M-F) Score	
	High	Low	High	Low	High	Low
Female Group						
Mother	10	4	5	9	8	6
Father	13	3	6	10	9	7
<i>p</i>	—		—		—	
Male group						
Mother	32	10	14	28	23	19
Father	9	9	10	8	5	13
<i>p</i>	.05		.10		.05	

TABLE 3
COMPARISON OF PERCEIVED SIMILARITY AND
INVOLVEMENT MEASURES OF
IDENTIFICATION
(Chi-square analysis)

Interview "like" Item	TAT							
	Mother Card		Father Card		(M-F) Score		Interview "child" Item	
	High	Low	High	Low	High	Low	Mo	Fa
Female group								
Mother	13	3	6	10	10	6	11	3
Father	8	3	5	6	5	6	2	8
<i>p</i>	—		—		—		.01	
Male group								
Mother	8	14	8	14	13	9	15	7
Father	16	17	20	13	12	21	23	10
<i>p</i>	—		.07		.08		—	

Hypothesis 3

The final analysis pertains to the prediction that measures of the perceived similarity aspect of identification (interview "like" item and semantic differential) are positively associated with measures of the involvement aspect of identification (TAT and interview "child"). Of the two perceived similarity measures of identification, only the interview "like" item was associated significantly with one of the involvement measures. From Table 3 we note that in the case of women, the interview "like" item relates significantly to the interview "child" item. For men, the interview "like" item relates in the predicted direction of identification, the Father card and the (M-F) score. The semantic differential did not relate sig-

nificantly to any of the involvement measures of identification.

DISCUSSION

Our finding that of the three semantic differential scores only the D score relates significantly for both sexes to the direct perceived similarity measure has implications for studies which utilize indirect measures of identification, such as rating scales or inventories. It suggests that a *comparative* score, comparing S's similarity to one parent in contrast to the other, may be a more fruitful measure of the perceived similarity aspect of identification than a score reflecting S's degree of similarity to either parent alone. Alternatively, if only one parent is used in deriving an identification score based on perceived similarity, the score based on the *same-sex* parent may be more desirable.

While our efforts to relate direct and fantasy measures of involvement met with some success in the case of men (Table 2), for women none of the results relating direct and fantasy measures approached significance. In part, we believe this is because the stimulus material used to elicit fantasy was less adequate for our purposes in the case of women than of men. For males, the two TAT cards depicted a "son" close in age to our Ss, while for females, the Mother card contained a "daughter" who was considerably younger than our female Ss. On the Father card for females, the stimulus material elicited an overly uniform theme in which 70% of the women told stories of a male figure expressing strong anger toward the female figure.

It is possible that the lack of significant findings for women on the TAT is related to the observation often made that parental identifications of women are less consistent than those of men. Sex differences in consistency of parental identification may be analyzed in terms of changes in scores from direct to indirect (or fantasy) assessment levels. In relation to the two aspects of identification we have measured, the findings suggest that, although men are more consistent than women in the involvement aspect of identification (Table 2), both sexes tend to be consistent in the perceived similarity aspect of identification (Table 1).

On the interview "like" and semantic differential D score, 60% and 51% of the men, respectively, perceived themselves as more like the father (Table 1). For women, 69% and 37%, on these two measures, perceived themselves as more like the mother. More detailed analysis of these shifts in identification between assessment levels reveals further differences between men and women. Of the 22 men who perceived themselves as more like the mother on the interview "like," seven (about one-third) shifted identification to the father on the D score. Conversely, of the 33 men with father identification on the interview "like," 12 (about one-third) shifted identification to the mother on the D score. Women showed a different pattern of shifts. Of the 16 women who stated that they were more like the mother on the interview, seven shifted identification to the father on indirect assessment. Of the 11 women with father identification on the interview "like," only one shifted to the mother on the D score. Thus, the relative inconsistency of the perceived similarity aspect of parental identification is different for both sexes. Women who shift, change from same-sex identification on direct assessment to opposite-sex identification on indirect assessment. Men who shift, change equally from one parent to the other.

Somewhat opposite results emerge on the involvement aspect of identification (Table 2). Here, women who shifted identification from the direct to the fantasy level tended to be equally divided in terms of changing from one parent to the other. For men, of those who changed, 79% stated they were closer to the mother on the interview but scored closer to the father on the TAT (M-F).

We conclude, then, that the relative inconsistency of parental identifications of men and women, as measured by a change in identification scores from the direct to indirect (or fantasy) assessment levels, is a function of two factors. First, consistency varies with the aspect of identification under study. Second, consistency may vary as a function of the particular parental identification established by direct assessment.

SUMMARY

This study was designed to investigate predicted relationships of three levels of direct-

ness of assessment (direct, indirect, and fantasy) to two aspects of parental identification, i.e., perceived similarity to a parent and degree of involvement with a parent. Perceived similarity was measured directly by an interview item ("Which parent do you feel you are more like?") and indirectly by the semantic differential. Involvement was measured directly by the interview ("Which parent were you closer to as a child?"), and by fantasy with two cards from the TAT or MAPS depicting a father-child or mother-child relationship. It was predicted that (a) perceived similarity measures would be positively related, (b) involvement measures would be positively related, and (c) since both perceived similarity and involvement are assumed to be components of identification, they should be positively associated with each other. Two groups of college students were used as Ss, consisting of 30 women in one group and 60 men in the other group. The major results are:

1. For both men and women, the direct (interview "like" item) and indirect (semantic differential) measures of perceived similarity are in general significantly related to each other (Hypothesis 1);

2. For men, the direct (interview "child" item) and fantasy (TAT) measures of involvement tend to be significantly related (Hypothesis 2);

3. For neither sex was the direct measure of perceived similarity (interview "like" item) and the fantasy measure of involvement (TAT) significantly associated, although trends in this direction were more apparent for men; for women, the direct measure of involvement (interview "child" item) was significantly related to the direct measure of perceived similarity (interview "like" item) (Hypothesis 3);

4. The problem of sex differences in consistency of parental identification is analyzed in terms of changes in identification scores from direct to indirect or fantasy assessment levels. It is concluded that both the aspect of identification measured and the specific parental identification made on direct assessment must be specified in drawing conclusions on this point.

REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- BEIER, E. G., & RATZEBURG, F. Parental identification of male and female college students. *J. abnorm. soc. Psychol.*, 1953, **48**, 569-572.
- CAVA, E. L., & RAUSH, H. L. Identification and adolescent boys' perception of the father. *J. abnorm. soc. Psychol.*, 1952, **47**, 855-856.
- FREUD, ANNA. *The ego and the mechanisms of defense*. New York: International Universities Press, 1946.
- KNIGHT, R. P. Introjection, projection and identification. *Psychoanal. Quart.*, 1940, **9**, 334-341.
- LAZOWICK, L. M. On the nature of identification. *J. abnorm. soc. Psychol.*, 1955, **51**, 175-183.
- LIVINGSTONE, D. The personality correlates of high and low identification with the father. Unpublished doctoral dissertation, Univer. of Houston, 1956.
- MOWRER, O. H. *Psychotherapy: Theory and research*. New York: Ronald, 1953.
- OSGOOD, C. E., SUCI, G. J., & TANNENBAUM, P. H. *The measurement of meaning*. Urbana: Univer. Illinois Press, 1957.
- PAYNE, D., & MUSSEN, P. H. Parent-child relations and father identification among adolescent boys. *J. abnorm. soc. Psychol.*, 1956, **52**, 358-362.
- SANFORD, R. N. The dynamics of identification. *Psychol. Rev.*, 1955, **62**, 106-118.
- SEWARD, J. P. Learning theory and identification. *J. genet. Psychol.*, 1954, **84**, 201-210.
- SOPCHAK, A. L. Parental identification and tendencies toward disorders as measured by the MMPI. *J. abnorm. soc. Psychol.*, 1952, **47**, 159-165.

Received December 23, 1957.

CASE REPORT

THE TREATMENT OF ANXIETY AND PHOBIC REACTIONS BY SYSTEMATIC DESENSITIZATION PSYCHOTHERAPY

STANLEY RACHMAN

University of the Witwatersrand, Johannesburg

THE aim of this paper is to present a detailed account of systematic desensitization psychotherapy in action. The present case was chosen for description because it clearly illustrates the principles and practice of this type of therapy and because of its intrinsic clinical interest.

The patient, Miss A. G., was a 24-year-old female teacher who complained of an inability to undergo injections of any kind. On those few occasions where she had been unable to avoid them, she had always fainted during or immediately after the injection. She requested therapy at the time she did because of an impending trip for which she would be required to have a yellow fever injection and smallpox vaccination. Her fear of injections was of long standing, dating back to either six or seven years of age. She experienced a moderate reaction (slight trembling and "butterflies in the stomach") when asked to imagine a person receiving an injection and ordinarily preferred not to talk about injections or related topics, e.g., visits to the dentist. She also complained of "a sexual problem" and a fear of using internal sanitary pads.

In addition to a symptom history, the patient was given the Willoughby Neurotic Tendency Inventory (Willoughby, 1934) and a form of the Incomplete Sentence Test during the first interview. Her Willoughby score was moderately high (44) and the I-S Test, which was only partially answered, revealed a slightly disturbed childhood, an exaggerated need for company and acceptance by other people, and vague fears about the future. Her relationships with her parents, with whom she lived, appeared to be unsatisfactory, but no serious difficulties seemed to be present. The general picture was one of mild insecurity.

In view of the nature and history of the chief symptom, it was decided to employ systematic desensitization psychotherapy and to attempt at the same time to relieve the feelings of insecurity by discussion and reassurance.

The systematic desensitization procedure was first described by Wolpe (1952b, 1954) in his attempt to develop techniques for the treatment of neuroses on the basis of reciprocal inhibition. The experimental basis (1948) for this type of therapy and the mechanisms and theory (1950,

1952a) underlying its formulation have been described elsewhere. The present paper is restricted to an account of systematic desensitization in practice.

TREATMENT

General Procedure

An inquiry is first conducted in desensitization therapy in order to ascertain which stimulus situations provoke anxiety in the patient. The patient is told that he can add to or modify this list at any time. The stimuli are then categorized by the therapist, and the patient is asked to rank the categories of stimuli in order from the most to the least disturbing. This ranked list of anxiety producing conditions is referred to as the hierarchy. In the present case, for example, one would refer to the "injection hierarchy" and the "sanitary pad hierarchy." Hierarchies typically contain from 5 to 25 items. The construction of the relevant hierarchies generally takes from one to three interviews, and the patient is concurrently given practice in hypnotic and relaxation procedures. Hypnosis is not an essential requirement, and, in those cases where the patient refuses to be hypnotized or requires prolonged practice, the procedure can be omitted and deep nonhypnotic relaxation employed instead.

When the hierarchies have been worked out, the subject is told which stimuli are to be presented in the individual session and is advised to signal with his hand if a stimulus presentation disturbs him unduly. This is an important instruction and is on no account omitted, because the arousing of anxiety during the session is sometimes extremely antitherapeutic. With most patients, it seems possible by observing facial expressions, bodily tension, respiration, and so forth, to perceive such disturbances before the patient actually signals. When such disturbances occur, the therapist immediately withdraws the stimulus and calms the patient. No session is concluded when a disturbance occurs, but before rousing the patient, the therapist presents an "easy" stimulus which has already been successfully overcome.

When the preliminary instructions have been given, the patient is relaxed (hypnotically or

otherwise) and then told to visualize the various stimuli, e.g., "Picture a hospital in the distance. . . . Now stop picturing that and go on relaxing." Each visualization of this kind is referred to as a "presentation." Each stimulus is visualized for five to ten seconds, and from two to four different items are presented each session. Each item is generally presented twice. When the requisite number of stimuli have been presented, the patient is slowly roused and then asked for a report on his reactions. If the items were visualized vividly and without undue disturbance, the therapist then proceeds to the next stimuli in the following session. The items lowest in the hierarchy (i.e., the least disturbing ones) are introduced first, and the therapist proceeds slowly up the list, depending on the progress achieved and the patient's reactions. In this way it is possible for the patient eventually to imagine formerly noxious stimuli without any anxiety whatever. This ability to imagine the noxious stimulus with tranquillity then transfers to the real-life situation.

Interview No. 2. A discussion of A. G.'s responses to the Incomplete Sentence Test revealed that some 18 months earlier, she had unsuccessfully attempted to use an internal sanitary pad. On the first attempt she fainted after the pad had been inserted. She left it in when she had regained consciousness, but it felt uncomfortable, and she was anxious lest she would not be able to extract it. The second attempt, a day later, was unsuccessful. She only managed to place the pad half way in after much effort. Her hand seemed to be stiff and almost paralysed, and she was perspiring and trembling. This failure left her very upset. A third attempt on the following day was also unsuccessful, and the emotional upset was repeated. Since that time she had used external pads exclusively. The patient also reported that she experienced pain during intercourse. This sexual difficulty had been present from her first experience of sex and was consistent. Her menstrual periods were regular, and she rarely experienced pain or other difficulties.

During this second interview the patient was given her first lesson in relaxation to which she responded very well and achieved a calm state within 15 minutes. She was instructed to practice relaxation for 15 minutes every day.

Interview No. 3. As the reaction to internal sanitary pads was judged to be less disturbing than that of injections, it was decided to commence with this problem. The following hierarchy was constructed for the patient progressively to imagine: box of Tampax, an opened pad, holding a pad, holding a pad next to the vagina, seeing someone else using a Tampax, placing a pad at the vaginal entrance, sliding it in slowly, completely

inserted. The patient was given a second lesson in relaxation and again responded well.

Interview No. 4. After a discussion about her ambitions and future plans, the patient was relaxed and asked to visualize each of the first three items in the hierarchy twice (i.e., box of pads, opened box of pads, holding a pad). She reported afterwards that the images had been slightly disturbing but not very vivid.

Interview No. 5. In the fifth session, A. G. revealed the history of her protracted relationship with the man she was in love with. She expressed doubts about the wisdom of marrying him and also about his attitude towards her. After some discussion, her feelings and motives became more lucid and she experienced some relief. Items 3, 4, and 5 in the anxiety hierarchy were successfully presented but again were not very vividly perceived.

Interview No. 6. The patient reported that she "felt better" since the last interview and had been practicing relaxation regularly. In view of her inability to visualize the hierarchy items vividly, she was given some instruction and practice to remedy this. She was then relaxed and Items 5 and 6 were presented three times each. The images were reported as slightly disturbing and a bit more clear.

Interview No. 7. A. G. said that she had been feeling tense for the past two days, but no incident or other cause for this upset could be located. She had seen a bullfight film after the development of this tension and reported that the bullfighting and darting of the bulls, previously upsetting to her, had left her unaffected. Desensitization of Items 6 and 7 (placing a pad at vaginal entrance, sliding it in slowly) was proceeded with. Each item was presented three times, and A. G. reported vivid images and a slight disturbance.

Interview No. 8. The patient's menstrual flow had started the previous afternoon, and she decided to try an internal sanitary pad. She reported, "I had a bath, felt a bit nervous, and then tried to insert a Tampax in the bathroom. I started perspiring, felt very hot, and got terribly upset. I tried to relax myself, then managed to put the tip of the pad in. Just then I had a 'thing' (near panic, almost fainted). I should have stopped but could not. I pushed and pushed and got terribly upset. I felt scared and started crying. Then C. (a friend) came in and tried to help me. I could not do it and eventually gave up. I felt dizzy and weak and was extremely upset."

The patient was given deep relaxation for 15 minutes and then the last two items were presented twice each (sliding pad in, pad fully inserted). The last item was more than usually disturbing. She was then given another five minutes

of relaxation and told to attempt an actual insertion in an adjoining room in the presence of a friend. This she managed to do with some difficulty after three minutes. When she had succeeded, however, she felt extremely pleased and had no dizzy spells or feelings of weakness.

She was instructed to insert a pad by herself that night after relaxing for 10 minutes and also to relax for a further 10 minutes after insertion.

Interview No. 9. A. G. reported that she had successfully inserted the sanitary pads on four occasions and had experienced two failures, neither of which had upset her very much. She said that she was now able to insert them more quickly and with little or no pain.

During this interview, an injection hierarchy was constructed. The items were (in ascending order) seeing a hypodermic syringe, holding a syringe, filling a syringe, seeing a cinema slide of a person receiving an injection, seeing a bull receive an injection, seeing a dog receive an injection, another person being injected, being injected at home, being injected at the district surgeon's rooms. The first three items in this hierarchy were then presented twice each. They were visualised vividly and were not disturbing.

Interviews 10 and 11. These sessions were mainly devoted to systematic desensitization. By this time, all the items up to and including No. 8 (seeing another person being injected) had been successfully presented.

Interview No. 12. The patient having received a jolt in her love relationship, this session was restricted to a sort of nondirective, cathartic discussion. No desensitizing was undertaken because of A. G.'s depressed mood and obvious desire to "just talk."

Interview No. 13. The cathartic process of the last interview was continued, and the patient was subsequently relaxed. Item 7 (dog being injected) was given once, Item 8 (person being injected) three times, and Item 9 (A. G. being injected at home) presented once very briefly. This last item caused some disturbance and was therefore not repeated in this session.

Interview No. 14. In anticipation of the patient's menstrual period, the sanitary pad image was reinforced under deep relaxation three times. Item 8 of the injection hierarchy (person being injected) was also visualised three times.

In the meantime, various difficulties regarding A. G.'s projected trip had arisen and were fully discussed.

Interview No. 15. Since the onset of her menstrual flow two days earlier, A. G. had successfully inserted sanitary pads on three out of four attempts. She was put under deep relaxation and then instructed to insert a pad while alone in the

relaxation room. She managed perfectly in very rapid time with no disturbance whatsoever.

She reported that she had attempted sexual intercourse a week earlier but had been forced to give up because of the pain and anxious feelings engendered. She added hesitatingly that she was worried lest she had some physical defect which would always prevent her from experiencing anything but pain in sexual activities. Analysis of the unsuccessful sexual attempt indicated that a contributing factor was inadequate foreplay. The patient was given information and advice about loveplay and told to relax fully before lovemaking. It was also suggested to her that there would probably be a spontaneous transfer of relaxation effect, and, therefore, success in sex was likely once the other two anxiety areas had been desensitized.

Interview No. 16. The menstrual flow had ceased and the "score" for the month was that six out of seven attempts at insertion of sanitary pads were successful. A. G.'s realization of this success had a marked beneficial effect, and she was quite elated about a long behavior difficulty which had at last been remedied.

The patient was then desensitized to Items 8 (person being injected) and 9 (self being injected at home). Each item was presented three times, and the patient reported no disturbance but inadequate visualization of the images.

Interview No. 17. A. G. stated that her trip was almost certainly cancelled and expressed considerable disappointment. She was also experiencing further uneasiness about her love relationship as a result of some action on the part of her partner which had given her cause to doubt his positive, affectionate feelings towards her.

Desensitization of the last two items in the hierarchy was carried out successfully. Each image (self being injected at home, being injected at district surgeon's rooms) was presented three times. They were vividly pictured and caused little disturbance, the patient being very deeply relaxed and extremely calm.

Interview No. 18. A. G. reported that she had experienced sexual intercourse two days previously. For the first time in her life, it had been completely free of even the slightest pain. She had felt slightly anxious but had managed to control this reaction and to indulge in pre- and post-coital loveplay unhindered. The fear of some physical defect had disappeared entirely, and she felt reassured about her sexual adequacy.

The last two items on the injection list were presented again, three times each. They were vividly imagined and caused no disturbance.

Interview No. 19. In reply to A. G.'s queries, information regarding the etiology of neurotic

behavior was supplied, and her own case was then discussed in some detail. She was told that phobias such as hers develop out of painful experiences and that despite their apparent senselessness they nevertheless persist. The usual reason for the persistence of the phobia, she was informed, is that it produces an avoidance of the painful situation. Because of her fear of injections, she had successfully managed to avoid having any form of injection for many years. Both the fear and avoidance reactions were reinforced over the years because they were never followed by pain and were in this way satisfying patterns of behavior.

The patient was desensitized to injections at the district surgeon's (three times) and to the sanitary pad situation (three times) in anticipation of the menstrual period again. Both items were seen vividly and without disturbance.

Interview No. 20. The patient reported that sexual intercourse had again been successful. She had experienced minimal anxiety and no pain. She also felt a change in her attitude to injections: "They no longer seem to bother me when I think about them."

After a final three presentations of the district surgeon situation under relaxation, it was agreed to test her reaction to injections. It was accordingly arranged that she would receive an injection of chemically pure water at the next interview.

Interview No. 21. After 10 minutes of deep relaxation, A. G. received an intramuscular water injection in the left arm. She experienced considerable pain but did not faint despite a strong feeling of "butterflies" and excessive sweating. She was relaxed for a further 15 minutes, but her arm continued to ache and she felt "shaken up."

The sanitary pads were now being used at will and provoked no anxiety or other untoward reactions.

Interview No. 22. Four days after the water injection, A. G. received her yellow fever injection at the district surgeon's rooms. She relaxed on a couch before and after the injection and experienced no disturbance despite her marked fear just prior to the event. She was extremely glad about this success and reported feeling "a lot better all round." As A. G. had been desensitized to the full hierarchies of noxious stimuli and her behavior difficulties overcome, this interview brought to an end the formal desensitization treatment. The patient was instructed to return for follow-up interviews at two-month intervals unless, of course, she felt the need to return before the stated time had elapsed.

Six weeks later, A. G. reported that she had received a smallpox vaccination in the interim and had experienced no ill effects although she

had been apprehensive for a while prior to her visit to the district surgeon's rooms. All the other improvements effected during therapy had been maintained.

Six weeks later, A. G. reported feeling well and "over her troubles." She was given the Willoughby Neurotic Tendency Scale again. Her score was 26, a decrease of 18 points since the first interview. Therapy was terminated as A. G. seemed improved in terms of Knight's criteria (1941) of symptom improvement, improved adjustment and pleasure in sex, and increased stress tolerance. The other two criteria of improved interpersonal relationships and increased productivity were not relevant to the case as A. G.'s behavior in these areas had never been disrupted.

DISCUSSION

The 22 therapeutic interviews were spread over a period of three months, averaging two per week. Perhaps the most striking feature of the case was its smooth progression. The therapeutic program proceeded from interview to interview in a regular, predictable way, with behavior changes following therapy in a manner almost perfectly consistent with theoretical expectations.

At only two points was therapy ever threatened with disruption. In Interview No. 12, A. G. arrived for her appointment depressed and perturbed about her love relationship. This mood prevented a desensitization session, and had it not cleared as rapidly as it did, it could have delayed further progress for weeks or even months. The second difficulty occurred towards the end of therapy as a result of the therapist's avoidable error in using water in the trial injection (Interview No. 21). The injection of water is ordinarily painful. The pain experienced by A. G. on this occasion fortunately did not reinforce her fear of injections unduly. She still underwent the yellow fever injection successfully a few days later. These two obstacles to progress have of course no general lesson or application other than the observation that therapeutic planning and procedure must be carried forward with considerable caution on the part of the psychotherapist.

It seems likely that the improvement in A. G.'s sexual performance resulted from a transfer of her progress in relaxation and in the sanitary pad procedure. Before the anxiety associated with the insertion of sanitary pads had been fully overcome, she unsuccessfully attempted to have intercourse (in the period between Interviews 13 and 14). By Interview 16, she was managing the sanitary pad insertions with very little anxiety and, soon after the following interview, had her first painless experience of sexual intercourse. Prior to the inhi-

bition of the sanitary pad anxiety she had been unable to have normal sexual intercourse. Spontaneous recovery of sexual functioning accompanying general psychological improvement is not unusual, but, in the present case, the transfer effect occurred rather early. The probable reason for this occurrence was the close similarity in this case between the symptom under treatment at the time (internal sanitary pads) and the pain experienced during sex.

The present case study indicates practically the point stated elsewhere (Lazarus & Rachman, 1957) that "while a knowledge of the causative process and genesis of the individual neurosis can be of considerable value in therapy, improvement can nevertheless be obtained in many cases without such knowledge" (Rachman, 1958). In the present case, no certain "cause" could be found for the development of the anxiety and phobic reactions. A. G. said that when she was a young child she had experienced some painful injections. This may well account for her neurotic reaction to injections, but the sanitary pad and sexual difficulties were never adequately traced backwards. The possibility that these problems arose out of the fear of injections as a prototypical fear of penetration seems farfetched. The reverse explanation, that all three "reactions to penetration" were of a sexual nature, is also not supported by the evidence. The phobic response to injections antedated the sexual difficulties by 12 years, and the penetration analogy can only be assumed to be relevant if one stretches the point. In any event, A. G. was assisted without either her or the therapist's discovering adequate reasons for the development of her behavior problems.

Although the technique of desensitization is of recent origin, successes with other types of psychological disturbance have been reported. Of the 122 cases reported by Wolpe in 1954, 72 were classified as anxiety states, 9 hysterics, 10 depressives, 11

obsessional-compulsives, 3 neurasthenics, and 17 mixed. He obtained cures or marked improvements in 110 cases (90%). Although a fuller discussion of the applicability of this therapy is provided elsewhere (Lazarus & Rachman, 1957), it seems safe to predict that phobias are particularly amenable to the desensitization technique.

SUMMARY

The treatment by systematic desensitization psychotherapy of a 24-year-old female patient suffering from anxiety and phobic reactions is described in detail. Several points arising out of the case history, including the development of the neurotic behavior and the "spontaneous" recovery of sexual adequacy, are discussed.

REFERENCES

- KNIGHT, R. P. Evaluation of the results of psychoanalytic therapy. *Amer. J. Psychiat.*, 1941, **98**, 434.
- LAZARUS, A. A., & RACHMAN, S. The use of systematic desensitization in psychotherapy. *South African Med. J.*, 1957, **32**, 934-937.
- RACHMAN, S. Objective Psychotherapy: Some theoretical considerations. *South African Med. J.*, 1958, **33**, 19-21.
- WILLOUGHBY, R. R. Norms for the Clarke-Thurstone Inventory. *J. soc. Psychol.*, 1934, **5**, 91-97.
- WOLPE, J. An approach to the problem of neurosis based on the conditioned response. Unpublished doctoral dissertation, Univer. of Witwatersrand, 1948.
- WOLPE, J. Genesis of neurosis: An objective account. *South African Med. J.*, 1950, **24**, 613-617.
- WOLPE, J. Experimental neurosis as learned behavior. *Brit. J. Psychol.*, 1952a, **43**, 243.
- WOLPE, J. Objective psychotherapy of the neurosis. *South African Med. J.*, 1952b, **26**, 825-829.
- WOLPE, J. Reciprocal inhibition as the main basis of psychotherapeutic effects. *A.M.A. Arch. Neurol. Psychiat.*, 1954, **72**, 205-226.

Received February 11, 1958.

CRITIQUE AND NOTES

MEASURING INTRA-INDIVIDUAL VARIABILITY WITHIN ONE TESTING¹

WALTER J. RAINE

Menninger Foundation

AND JOHN R. HILLS

Regents of the University System of Georgia

RECENTLY the concept of intra-individual response variability has been receiving considerable attention by psychologists (Fiske, 1957a; Fiske, 1957b; Fiske & Rice, 1955; Klein, 1954; Smith & Klein, 1936). Fiske and Rice (1955) in a comprehensive article on the subject discussed several concepts of variability. One concept depends on discrepancies between responses to homogeneous stimuli, i.e., to stimuli so objectively similar that the same response is expected. Another depends on discrepancies when identical stimuli are presented on separate but very similar occasions. It remains to be seen whether these two kinds of intra-individual variability are related or, perhaps, psychologically the same.

In order to examine this question it was decided to incorporate a study of intra-individual variability into a larger study of personality instruments which were administered to entering law school students.

The Problem. Another way of phrasing the question posed in this research is to ask whether one can improve upon the technique Fiske has used for measuring variability (repeated testing with the same items) in order to make more feasible the measurement of intra-individual response variability using large groups of people scattered in many testing centers over the entire country. For our purpose, the time available for testing was limited to three hours, and it was proposed to administer a complex battery of tests to each individual. Multiple administration of the same test seemed likely to generate hostility among subjects; it also would add time to test administration. If measures of intra-individual variability are to be useful in the guidance of large numbers of people, some way must be found to shorten the time and to avoid the test-retest situation.

¹ This research is part of a larger study conducted by Educational Testing Service with the cooperation of several member schools of the Law School Admission Test Policy Committee. The project is under the direction of John R. Hills. During this study the authors were employees of the Educational Testing Service, Princeton, N. J.

The Hypothesis. The hypothesis was advanced that the test-retest situation could be duplicated within the context of one test by using psychologically equivalent items. That is, there should be a significant positive correlation between variability scores obtained in repeated testing with the same items and scores obtained by comparing responses made to psychologically equivalent items within a single test. It was felt that the Sentence Completion Test lent itself well to the use of items that were psychologically equivalent in terms of projection. It was hypothesized that stems using the first person (I think women . . .) would be psychologically equivalent for men to stems using the masculine third person (He thought women . . .). (All subjects were males.) In this way stems already present in the Sentence Completion Test could be changed superficially, but in effect the subject would be responding to those items twice.² The study reported here tests this hypothesis.

Procedure. The Sentence Completion Test used by Fiske (1957b), augmented by 15 more items hoped to be psychologically equivalent to 15 items within the original 25, was administered twice to a group of 74 entering law school students at a

² The assumption of psychological equivalence in the Sentence Completion Test when using first versus third person stems may seem unwarranted in the light of two previously reported studies (Hanfmann & Getzels, 1953; Sachs, 1949). Both of these studies hypothesized that the responses to first person stems would differ from the responses to third person stems. Both found support for this hypothesis. We, however, are expecting to find similarities between responses to first and third person stems. It is entirely possible that the two kinds of responses can be both similar and different. In that case, the correlation between variability scores based on responses to identical items and variability scores based on "equivalent" items will be significantly different from zero but also significantly different from 1.00 when corrected for attenuation. This is precisely what occurred in our data. In addition to the above, when comparing our study with the two earlier studies, one immediately becomes aware of the important differences between the three investigations in terms of procedures, subjects, and item content. These factors alone drastically limit the relevance of their conclusions to our initial hypothesis.

large eastern state university. The first administration took place at the beginning of the test battery and the second took place three hours later at the end. The instructions were identical in both cases.

Scoring. The Sentence Completion Test is scored in this study as it was scored by Fiske,³ on the basis of similarity of the repeat responses to the original. A stringent criterion of similarity was adopted. The number of such similar responses is the test-retest measure of intra-individual variability. A high similarity score is, therefore, equivalent to a low variability score.

In our situation there are a number of possible scores. Score 1 is the number of similar responses given by an individual to the same item for the first 25 items of the augmented Sentence Completion Test in the two administrations. (This set of items is Fiske's original Sentence Completion Test.)

Score 2 is the number of similar responses given by an individual to the same item for the last 15 items of the augmented Sentence Completion Test in the two administrations.

Score 3 is the number of similar responses given by an individual to the same item for the total 40 items of the augmented Sentence Completion Test in the two administrations and is the sum of Score 1 plus Score 2.

Score 4 is the number of similar responses given by an individual to 16 pairs of psychologically equivalent items within the augmented Sentence Completion Test during the first administration. The criteria of similarity for this score were analogous to the criteria for the previous three scores. (Although the test was augmented by adding 15 psychologically equivalent items, the original set of 25 items was found to contain a pair that also seemed equivalent in these terms.)

Analysis. The distribution of Score 1 for the 56 students who completed the first 25 stems in the two administrations is rather flat but quite symmetrical with mean of 9.8, median of 10, mode of 11, and standard deviation of 5.8. The range of scores is from 0 to 24. The small sample prohibits definitive conclusions, but it can be said that test-retest intra-individual response variability does appear in one testing session. The effect seems to be distributed rather evenly between scores of 0 and 19.

For the purpose of further analysis it was decided to use the cases that met the following criteria:

1. At least 20 out of 25 completed pairs of responses on the first 25 stems.
2. At least 31 completed pairs of responses on the full 40 stems.
3. At least 10 out of the last 15 pairs of stems completed.

This resulted in an N of 59. The Pearson product-moment correlation was computed for Score 1 with Score 2, and Score 2 with Score 3, and Score 1 with Score 4. Reliability estimates were computed (Kuder-Richardson No. 20) for all four scores and the appropriate corrections for attenuation were made.

Results. The internal consistency reliability of Score 1 was estimated to be .85. For Score 2 it is .80, for Score 3 it is .90, and for Score 4, .52. The correlation between Score 1 and Score 2 was .81, which corrected for attenuation in both scales, is .98. This finding indicates that Score 3 (the sum of Scores 1 and 2) is measuring the same variables as Score 1. Score 3, as we have seen, is superior in reliability, having 60% more items.

The correlation of Score 1 with Score 4 was .55 ($p < .01$). This is the correlation between the number of similar responses in the test-retest situation for the first 25 stems (Fiske's original situation) and the score obtained from comparing the responses to the 16 pairs of psychologically equivalent items within the first administration of the Sentence Completion Test. Corrected for attenuation in both scales, this correlation is .83.

Score 4 correlated .61 with Score 3 ($p < .01$). This is the correlation between the number of similar responses made on the full 40 items in the test-retest situation and the score for the 16 pairs of psychologically similar items within the first administration. Corrected for attenuation it is .90, approximately in agreement with what we would expect from the superior reliability of Score 3 as compared with Score 1.

Conclusions. The two correlations of .55 and .61 are representative of the ability of the particular psychologically equivalent items used in this test to reproduce the test-retest variability. The corrected correlations of .83 and .90 indicate that the effect of test-retest variability can be duplicated to a large extent within the context of one administration, and that with more items of the "equivalent" type a satisfactorily reliable measure of this effect can be obtained without resorting to several administrations. Since the extra items were added without any prior testing of their equivalence, it may be presumed that with further study better items could be written that would result in higher correlations. The results are interpreted as being promising, rather than definitive. It is to be hoped that research on intra-individual variability will

³ Fiske transmitted his criteria for scoring to the authors in a personal communication. The authors are most appreciative of D. W. Fiske's interest, cooperation, and critical comments on this manuscript.

be continued, and that this demonstration of the possibility of achieving test-retest variability within the context of one administration will materially aid such research.

REFERENCES

- FISKE, D. W. The constraints on intra-individual variability in test responses. *Educ. psychol. Measmt.* 1957a, 17, 317-337.
- FISKE, D. W. An intensive study of variability scores. *Educ. psychol. Measmt.* 1957b, 17, 453-465.
- FISKE, D. W., & RICE, LAURA. Intra-individual response variability. *Psychol. Bull.*, 1955, 52, 217-250.

- HANFMANN, EUGENIA, & GETZELS, J. W. Studies of the Sentence Completion Test. *J. proj. Tech.*, 1953, 17, 280-294.
- KLEIN, G. S. Need and regulation. In M. R. Jones (Ed.), *Nebraska symposium on motivation: 1954*. Lincoln, Nebr.: Univer. Nebraska Press, 1954.
- SACKS, J. M. The relative effect upon projective responses of stimuli referring to the subject and of stimuli referring to other persons. *J. consult. Psychol.*, 1949, 13, 12-20.
- SMITH, G. J., & KLEIN, G. S. Cognitive controls in serial behavior. *J. Pers.*, 1936, 22, 181-188.

Received November 18, 1957.

THE RELIABILITY OF THE ACHIEVEMENT MOTIVE

ROBERT C. BIRNEY

Amherst College

ALTHOUGH the use of projective techniques for the measurement of the achievement motive (n Ach) has been well established, the reliability of the measure is not high (McClelland, Atkinson, Clark, & Lowell, 1953). Product-moment coefficients of equivalence-stability for three- to five-week intervals have ranged from +.22 (Lowell, 1950) to +.64 (Morgan, 1953). This is a report of coefficients of equivalence-stability obtained through the administration to the same Ss of pictures in parallel forms, in October 1956, February 1957, and April 1957.

Previous research (Birney, 1957) has indicated that when Ss are asked to rate pictures on the attributes that comprise the n Ach scoring system, a sufficiently high correlation (τ_{aa} , +.39) is obtained to permit tentative assignment of the pictures to parallel batteries. In addition, the percentage of Ss writing achievement stories, i.e., stories scored for achievement imagery (Ai), in response to each picture can be used for picture selection. Both ratings and data on incidence of achievement imagery (Ai) gathered by Haber and Alpert¹ were used in selecting three batteries of four pictures each for administration.

Sixty college males in the sophomore class at Amherst College were chosen at random and invited to participate in the study. Test sessions scheduled for October and February were chosen for their primarily "academic" emphasis; while the April session preceded by one week the most im-

portant social event of the semester. Forty-six Ss appeared for the October session. The measures of achievement motive were administered² in standard fashion (McClelland et al., 1953). Following the story-writing, the Ss were divided into two groups for further testing.³ The last task of each session required the Ss to fill out a short questionnaire about the most important features of their current campus life. For the February session, 40 of the original 46 Ss returned, and the same student E administered the second form of the test. Finally, in April, 24 Ss answered the third summons.

The three batteries of pictures have been called parallel because they were selected on the basis of prior use and rating as having matched stimulus qualities. Table 1 shows that Forms A and B are psychometrically parallel having equal means and variances, while the Form C mean is reliably lower than the other two forms.

Table 2 presents the product-moment correlations between forms. Included are the coefficients obtained between the scores gathered in the 1956-1957 test sessions and scores obtained in August 1955 from these same Ss, using another set of six pictures.⁴ The self-rating material supported the initial assumption that academic concern pre-

² The author is indebted to James Allen for his assistance.

³ These data are not relevant to the present discussion.

⁴ The pictures used were #2, inventors; #8 boy in checkered shirt; #1, father-son, (TAT 7BM); two students conversing; five men in a club; and a man in a doorway.

¹ The author wishes to express his appreciation to Ralph Haber and Thomas Alpert for making available their data and materials gathered at Stanford University.

TABLE 1
COMPARISON OF PICTURE FORMS

Form A (N = 46)		Form B (N = 40)		Form C (N = 26)	
Picture ^a	Mn Ach	Picture ^b	Mn Ach	Picture ^a	Mn Ach
a	1.65	2	1.40	e	.65
b	-.13	11	.48	f	1.15
c	-.33	34	-.05	g	-.65
d	1.46	46	1.05	h	-.46
Mean	2.66		2.88		+.69*
SD	4.84		4.30		3.59

^a Pictures supplied by Haber and Alpert are described as follows: (a) three men and still; (b) hand, cap, gown, diploma; (c) man sleeping under paper; (d) home carpenter's fantasy; (e) chemist and test tube; (f) man sitting by vat; (g) man and tractor; (h) little Korean girl.

^b Picture code from McClelland's list of pictures: (2) inventors; (11) engineer with apparatus; (34) bar scene; (46) skier.

* Form C mean is significantly lower than Forms A and B at $p < .01$.

dominated in February and October, and social concern in April. These considerations point to the October-February comparison as giving the most suitable estimate of score stability. These results indicate that Lowell's value of +.22 for a three-week interval falls near the midpoint of the range of values obtained here, and that better control procedures failed to produce higher reliabilities. Thus, it appears that the n Ach measure is highly situational in character as reflected by low coefficients of equivalence-stability, perhaps best repre-

TABLE 2
COEFFICIENTS OF EQUIVALENCE-STABILITY FOR n ACHIEVEMENT

	February 1957 (N = 40)	April 1957 (N = 26)	August 1955 (N = 46)
October 1956 (N = 46)	+.29	+.56*	+.15
February 1957		+.20	+.40*
April 1957			+.03

$p < .05$.

sented by the October-February value of +.29. Of additional interest is the fact that the coefficients with the August 1955 values do not cluster around zero.

REFERENCES

- BIRNEY, R. C. The cue characteristics of pictures and thematic content. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society: A method of assessment and study*. Princeton, N. J.: Van Nostrand, 1957.
- LOWELL, E. L. A methodological study of projectively measured achievement motivation. Unpublished master's thesis, Wesleyan Univer., 1950.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. Appleton-Century-Crofts, 1953.
- MORGAN, H. H. Measuring achievement motivation with "picture interpretations." *J. consult. Psychol.*, 1953, 17, 289-292.

Received November 22, 1957.

HYPOTHESES OF OPPOSITE SPEECH¹

JULIUS LAFFAL AND LANE AMEEN

Veterans Administration Hospital, West Haven, Connecticut

SUBSEQUENT to a paper by the present authors on opposite speech in a schizophrenic patient (Laffal, Lenkoski, & Ameen, 1956), two commentaries were published. Staats (1957) suggested possible etiological factors in opposite speech and some ways of dealing with it based on principles of reinforcement. Kaplan (1957) attempted to interpret the phenomenon in terms of Werner's developmental views. In this note, we wish to make clear our differences with these two authors. An excerpt

¹ This study is part of a continuing research on language distortions in schizophrenia, which is supported by USPHS Grant M-2020.

² From the Department of Psychiatry, School of Medicine, Yale University, and the U. S. VA Hospital, West Haven, Connecticut.

from an early recorded contact with the patient will illustrate what we have called opposite speech.

Dr.: . . . who invented the airplane?

Pt.: I do know.

Dr.: You mean, you don't know.

Pt.: I do know.

Dr.: You do know.

Pt.: Yes, I do.

Dr.: If you do know, can you tell me?

Pt.: If I do know, how can I tell you? I could.

Dr.: You could tell me.

Pt.: Yes, because I do know. I do know, I do know,

ah, who invented the airplane.

Dr.: Okay, if you do know who invented the airplane, tell me who invented the airplane.

Pt.: I can.

Dr.: You can.

Pt.: I sure could.

Dr.: You sure could. Okay, can you tell me now who invented the airplane?

Pt.: I do know.

Dr.: You do know.

Pt.: Yes, I know.

Dr.: That means that you have the answer. You have the answer to that question.

Pt.: Yes.

Dr.: Yes. All right, now can you tell me what the answer is?

Pt.: Who invented the airplane, I do know.

Dr.: What you mean to say is that you don't know.

Pt.: I do know. If I don't know, I, I, I, I wouldn't be able to tell you.

Dr.: You're not able to tell me, though, are you?

Pt.: Yes, I am, for I do know.

The developments in this case, which cover a period of nearly three years since the first report, are briefly as follows. The patient recovered from his psychosis while in the hospital and has been working and living productively outside of the hospital for the past 20 months. Immediately preceding his recovery, during a trial visit at home, he became extremely hostile and threatening toward his parents; even fearing for their lives, they returned the patient to the hospital with the help of the police. It was at this time that the opposite speech was first noted to be absent, and it was shortly after his return to the hospital that the patient's psychosis cleared up. The circumstances surrounding the disappearance of the opposite speech gave striking support to the first alternative of our prediction (Laffal et al., p. 413) "that if this man's speech syndrome were to clear we would either see the outbreak of aggression or find aggressive content in his fantasies." In the long retrospect, it is now apparent that the second alternative, relating to aggressive fantasy, may also be correct. Periodically in the therapy which has continued on a once-a-month basis since the patient left the hospital, he has brought up, with great hesitation and embarrassment, an anxiety-ridden, obsessive fantasy of being unexpectedly seized and subjected to shock treatment.

The points which Staats (1957) makes about the patient's syndrome are (a) the opposite speech elicits attention from others, and this serves as a reinforcement of the speech pattern; (b) lack of reinforcement, or withholding compliance with the patient's desires when he uses opposite speech, should weaken the opposite speech; and (c) anxiety accounts for the opposite speech, correct speech being anxiety-arousing for the patient because of its typical content. Staats (1957, p. 269) suggests as the source of the anxiety, "that the unhappy life situation of an adult schizophrenic probably elicits thought and speech which are not positive secondary reinforcers, but instead arouse anxiety."

The first of these points was considered in the original paper (Laffal et al., 1956, p. 410) in terms of the possibility that the opposite speech might be persisting because of certain secondary rewards, i.e., because of the interest it aroused. However, it was noted that relatively little time was devoted to the sheer exploration of the patient's speech syndrome itself, the emphasis in all contacts being on treatment. A strong argument that interest did not maintain the symptom is that the symptom disappeared while we continued to be interested in it.

Regarding Staats's second point: Despite demonstrations and remonstrances by doctors and others that he was reversing his speech, the patient continued to do so. Early in his hospitalization, the patient repeatedly demanded in opposite speech that he be discharged. The staff was far from complying with this request, but there was no evidence that the opposite speech changed on that account. It is true that no systematic effort was made to withhold reinforcement or to chide the patient whenever he used opposite speech. However, there is a paradox in Staats's argument in that to have used such a systematic approach would, of necessity, have involved giving the opposite speech the very attention which Staats believes supported it.

With respect to Staats's final point, anxiety was certainly a factor in the opposite speech. However, since the patient returned to his ordinary life situation and lost the symptom, it is doubtful that the syndrome is attributable to his "unhappy life situation."

Kaplan's (1957) treatment of the opposite speech syndrome is an attempt to conceive this phenomenon in terms of the developmental views of Werner. We differ with him on this central thesis (Kaplan, 1957 p. 390) that "opposite speech is presumed to arise because the content to which the linguistic signs refer is in such a global, undifferentiated state that the linguistic vehicles (seemingly discrete) really share the same global referent." In our original paper we pointed out that opposite speech consists (Laffal, et al., 1956, p. 409) "basically in the use of 'yes' by the patient when he means 'no,' and vice versa. . . it includes the interchange of 'right' and 'wrong,' the interchange of 'do' and 'don't,' and occasionally the interchange of such opposites as 'something' and 'nothing' and 'like' and 'hate.'" We stressed the disparity between the conscious intention of the patient and the verbalizations he used to communicate the intention. Kaplan (1957, p. 390), in talking about "a lack of differentiation in the use of affirmative and negative forms of judgment, e.g., 'I know,' 'I don't know,' etc.," is not talking about opposite speech as we have described it, but

about a phenomenon more in keeping with his developmental notions, namely, an early state of language development in which some area of reference is not sufficiently discriminated by the subject to permit distinctive verbal labeling either of parts or even of extremes of the area.

In the opposite speech of the patient under consideration, it is not that the referent or the speech is undifferentiated, but that the language operations are appropriate to the opposite of the referent. The patient intends something quite succinct and differentiated, but uses opposite speech. Thus, if one went through the interview excerpted above and interchanged *yes* and *no* and *do* and *don't*, one would have a nearly normal sounding conversation. Such a high degree of consistency of reversal would indicate that the referents remain distinct and differentiated, and also that opposite language operations are well discriminated by the patient.

The pathology lies in the fact that the language used by the patient is appropriate to the opposite of his referents or of what he intends.

The opposite speech phenomenon is open to many interpretations and may be looked at from many points of view. The interpretation originally offered (Laffal et al., 1956) that it was a function of fear of aggression has been consistent both with the circumstances surrounding the disappearance of the opposite speech, and with the subsequent content of the patient's productions in therapy.

REFERENCES

- KAPLAN, B. On the phenomena of "opposite speech." *J. abnorm. soc. Psychol.*, 1957, **55**, 389-393.
 LAFFAL, J., LENKOSKI, L. D., & AMEEN, L. "Opposite speech" in a schizophrenic patient. *J. abnorm. soc. Psychol.*, 1956, **52**, 409-413.
 STAATS, A. W. Learning theory and "opposite speech." *J. abnorm. soc. Psychol.*, 1957, **45**, 268-269.

AN ANALYSIS OF SOME FEATURES OF THE INTERVIEW WITH THE INTERACTION CHRONOGRAPH

SAMUEL B. GUZE AND IVAN N. MENSCH

Department of Psychiatry and Neurology, Washington University School of Medicine

THE USE of the Interaction Chronograph as an instrument for objectively assessing interview behavior has been described in a series of studies dealing with the method, its history and its place in psychological theory (Matarazzo, Saslow, & Matarazzo, 1956; Matarazzo, Saslow, & Guze, 1956; Saslow, Matarazzo, & Guze, 1955; Saslow, Matarazzo, Phillips, & Matarazzo, in press). These studies have involved the use of a psychiatric interview (Chapple, 1953) standardized in order to minimize the well-known effect of interviewer variability (Goldman-Eisler, 1952). The standardized interviews consist of five periods, the first of which is a baseline period of nonstressful interaction 10 minutes long in which the interviewer tries to make his utterances as nondirective as possible, approximately five seconds long each time, without either interrupting the subject or delaying his response more than a half second. A duration of 10 minutes was selected by Chapple (1953) on the basis of his experience, but there are not published data to show that 10 minutes are long enough to establish a stable baseline. The question of how representative the first 10 minutes of a subject's behavior under baseline conditions are of his behavior under such conditions over longer periods of time was not explored in the studies of Saslow et al. (1955; 1956), which fo-

cused on the reliability of the first 10 minutes of the standardized interview as measured by re-interviewing with the same standardized interview. The present study was undertaken to determine whether 10 minutes are sufficient to reach a stable pattern of patient communication under Period I conditions. If so, 10 minutes may be sufficient to use as a baseline when evaluating the effect of drugs, EST, and other stimuli upon such behavior.

Since the average length of the previous standardized interviews was 32 minutes (Matarazzo et al., 1956; Saslow et al., 1955), it was decided to study a new series of patients in 30-minute interviews, in which the interviewer maintained Period I conditions, and to compare the data of the three successive 10-minute portions. The patients were randomly selected white patients who were referred to the psychiatric clinic for the first time. The means, medians, standard deviations, and ranges of the patients' actions were computed for the three periods for each patient in the sample.

RESULTS

Table 1 summarizes the data of the 19 individual patients. Table 2 presents the summary data for the total samples. Patients' actions are defined as any communication, verbal or nonverbal, during

TABLE 1
PATIENTS' ACTIONS* DURING THIRTY-MINUTE
PSYCHIATRIC INTERVIEW, BY INDIVIDUAL

Pt.	N ₁ ^b	N ₂	N ₃	M ₁	Mdn ₁	M ₂	Mdn ₂	M ₃	Mdn ₃
1	7	8	7	104°	186	104	108	138	146
2	14	46	58	26	22	18	12	10	7
3	26	35	28	29	18	18	12	29	24
4	9	4	4	111	55	222	239	192	268
5	5	3	8	180	167	63	110	99	102
6	19	18	21	40	37	46	43	42	49
7	13	12	7	60	43	74	46	135	73
8	13	10	9	58	15	78	51	114	163
9	17	8	13	42	35	102	98	69	62
10	9	8	9	96	124	106	106	109	108
11	20	22	24	40	44	38	34	34	26
12	12	21	13	77	70	31	20	59	58
13	10	12	13	78	68	63	75	68	66
14	17	22	17	37	33	44	23	53	26
15	8	10	10	110	70	91	72	94	74
16	17	23	21	46	49	27	26	23	13
17	13	16	11	64	87	44	37	68	73
18	8	10	18	99	103	78	69	49	53
19	27	21	25	30	26	40	28	46	18

* Any communication, verbal or nonverbal, during period of observation.

^b Number of units in successive 10-minute periods.

^c These values represent units of time in hundredths of minutes and are rounded off to the nearest whole number.

TABLE 2
PATIENTS' ACTIONS* DURING THIRTY-MINUTE
PSYCHIATRIC INTERVIEWS, FOR TOTAL
SAMPLE (N 19)

Statistic	Period I ^b	Period II	Period III	Total
Total N units	264	309	316	889
Mean time/unit	63.90°	44.09	53.66	59.07
Median time/unit	65.84	63.63	74.10	58.00
SD time/unit	62.76	78.52	74.68	82.65
Range of N's/individual	5-27	3-46	4-58	3-58
Range of mean time/unit/individual	26.50-179.80	17.56-221.50	10.45-191.75	10.45-221.50
Range of median time/unit/individual	15-186	11.50-239	7-267.50	7-267.50

* Any communication, verbal or nonverbal, during period of observation.

^b Each period consists of 10 minutes.

^c These values represent units of time in hundredths of minutes.

the period of observation. The units of time are in hundredths of minutes. It can be seen that there are marked inter- and intra-individual variations in number and duration of actions. The number of actions range for the 19 Ss from 3 to 58 in the various 10-minute periods and from 16 to 118 in total number; while the durations of actions range from 10.45 to 221.50 units (6.5-133 seconds). The distribution of numbers of units of action approximated a normal curve, with slight skewing toward the upper end and one principal deviant score of 118. The intra-individual variations, however,

show marked skewing for some patients and but little for others, as indicated by the discrepancies between mean and median values. Thus, for the first 10 minutes of interaction, 10 of the 19 patients had essentially normally distributed patterns of action; for the second 10 minutes there were nine patients with similar action patterns; and for the third 10 minutes there were nine again; roughly half of the sample within any one period, but only two of whom, Patients 6 and 18, had action patterns essentially normally distributed throughout the three periods comprising the total interview. Of the 29 skewed distributions among the total of 57, 21 were positively skewed toward the lower end.

In addition to the skewed distributions for the 10-minute periods in this sample of 19 patients, further evidence of the nonnormal characteristics of these distributions of patients' actions in communication is provided in the extreme variability indicated by the large standard deviations. The ratios of range/SD (Snedecor, 1946) indicated that the distributions of 11 of the 19 patients met the probabilities of random sampling from a normal distributed population, and 8 did not.

Among the 11 distributions reported above, only 6, 7, 13, and 20 did not have significant skewing. It was thus apparent that one of the usual tests of reliability of sample means, the increase in reliability of the mean with decrease in standard error of the mean with decrease in standard error of the mean (Garrett, 1937) was not appropriate. Rather, a statistic such as the median would serve as a better estimate of the population central tendency parameter, where extreme scores would not influence the estimate.

Though there was great intra-individual variability across the three periods as indicated above, an analysis of the group data indicated that this variability did not represent purely random behavior. This would be expected, of course, since the interview was standardized, and this was indicated by correlations between medium scores, for the total group of 19 patients, of .58, .49, and .82, respectively, between Periods 1 and 2, 1 and 3, and 2 and 3.

DISCUSSION

The previous studies by Saslow, Matarazzo, and Guze (1955; 1956) established the high reliability (.73-.96) of the patient's interaction between two successive interviews by different interviewers when the values for the entire 30-minute interviews were compared. In their studies, the reliabilities were lower (.00-.91) when individual subperiods were compared.

It is clear that the variation within any single 30-minute interview between successive 10-minute intervals under baseline conditions is too great to

justify selecting 10 minutes as a baseline for comparison with other intervals during which different events are going on even though comparing the first 10 minutes of *one* interview with the first 10 minutes of a *second* interview with the same patient gives high reliability. These differences need further study. They indicate that an individual begins his interviews with a characteristic, reliable pattern but that this pattern does not remain stable if it is permitted to continue for a full 30 minutes, and that it does not vary in any regular fashion.

REFERENCES

- CHAPPLE, E. D. The standard experimental (stress) interview as used in interaction chronograph investigations. *Hum. Organization*, 1953, **12**, 23-32.
- GARRETT, H. E. Statistics in psychology and education. (2nd ed.) New York: Longmans-Green, 1937.
- GOLDMAN-EISLER, F. Individual differences between

interviewers and their effect on interviewees' conversational behavior. *J. ment. Sci.*, 1952, **98**, 660-670.

- MATARAZZO, J. D., SASLOW, G., & MATARAZZO, RUTH G. The interaction chronograph as an instrument for objective measurement of interaction patterns during interviews. *J. Psychol.*, 1956, **41**, 347-367.
- MATARAZZO, J. D., SASLOW, G., & GUZE, S. B. The stability of interaction chronograph patterns in psychiatric interviews: A replication. *J. consult. Psychol.*, 1956, **20**, 267-274.
- SASLOW, G., MATARAZZO, J. D., & GUZE, S. B. The stability of interaction chronograph patterns in psychiatric interviews. *J. consult. Psychol.*, 1955, **19**, 417-430.
- SASLOW, G., MATARAZZO, J. D., PHILLIPS, JEANNE S., & MATARAZZO, RUTH G. Test-retest stability of interaction patterns during interviews conducted one week apart. *J. abnorm. soc. Psychol.*, in press.
- SNEDECOR, G. W. *Statistical methods*. (4th ed.) Ames, Ia.: State College Press, 1946.

Received December 27, 1957.

POSSESSIVENESS AS A CHARACTERISTIC OF MOTHERS OF SCHIZOPHRENICS¹

HOWARD E. FREEMAN, OZZIE G. SIMMONS, AND BERNARD J. BERGEN

Harvard University

SEVERAL recent quantitative investigations, in addition to numerous clinical studies, have been directed at understanding the association between parent-child relations and schizophrenia (Dworin & Wyant, 1957; Freeman & Grayson, 1955). In one of these studies, Freeman and Grayson (1955) conclude that possessiveness is an attitude characteristic of mothers of schizophrenics "... which might be assumed to be of significance in the psychogenesis of schizophrenia" (p. 51). In this paper, we report the results of a study of the association between possessiveness of mothers and schizophrenia of sons, based upon items used by Freeman and Grayson.

METHOD

Freeman and Grayson utilized the Shoben Parent Attitude Survey (Shoben, 1949) to study maternal attitudes in schizophrenia. The 85 items of the survey are divided into three scales, one of which is described as measuring possessiveness, i.e., "... a tendency on

the part of a parent to baby the child, to emphasize unduly the affectional bonds between parents and child..." (Freeman & Grayson, 1955, p. 45). In the Freeman and Grayson study, the Shoben Survey was filled out by 50 mothers of schizophrenic patients and a control group of 50 mothers. The mothers of schizophrenics had sons between the ages of 20 and 35 who were hospitalized in a VA facility. The control group had children who had never required psychiatric attention. Mothers of schizophrenics were more possessive on almost all items in the scale, and the differences between mothers of the patients and control mothers were statistically significant on four of the individual items ($p < .05$). The mean difference was significant at the .02 level.

The present data were obtained as part of a study of the relationship between posthospital performance level of mental patients and characteristics of their female relatives (Freeman & Simmons, 1958). The four Shoben items that significantly discriminated mothers of schizophrenic patients and mothers of males not requiring psychiatric care were employed. These items failed to correlate with performance level (Freeman & Simmons, in press).

Female informants interviewed were all relatives, predominantly wives and mothers, of male patients who succeeded in remaining in the community after their latest release from a mental hospital sometime between November, 1954, and December, 1955. Every

¹ This research is being undertaken by the Community Health Project, under the direction of Ozzie G. Simmons. The project is sponsored by the Social Science Program at the Harvard School of Public Health and is supported by a grant (M 1627) from the National Institute of Mental Health.

male patient with the following characteristics was included in the potential study group: between 20 and 60 years of age, white, native born, living in the Boston metropolitan area at the time of release, hospitalized more than 45 days prior to release, not physically handicapped to the extent of being unemployable, not addicted to narcotics, and not primarily hospitalized for acute alcoholism. By diagnosis, all were psychotics with nonorganic disorders. Patients selected were last hospitalized in one of 13 hospitals in the Boston area. Of the 209 interviews attempted, 182 (88%) were completed. The informant group and selection procedures are fully described elsewhere (Freeman & Simmons, 1958).

In this research, unlike the Freeman and Grayson study, the items were part of an extensive interview schedule composed mostly of structured questions. This study can consequently be regarded more properly as a comparative investigation rather than as a replication. Not only are the items embedded in a different instrument, but the responses were obtained as part of an interview rather than as answers to a questionnaire.

RESULTS

The responses of the mothers of schizophrenic patients were compared with those of four other groups: (a) mothers of patients with other nonorganic disorders (mostly manic-depressives); (b) wives of schizophrenic patients; (c) wives of patients with other nonorganic disorders, and (d) other female relatives of schizophrenic patients (mostly their sisters). As Freeman and Grayson note, the original weights may not be useful for the study of relatives of adult patients because they were developed from the responses of a sample of mothers of disturbed children. The conclusion remains the same, however, whether or not the mean weights or the distributions of responses along the four-point scales are the basis of the analysis. Among the informants, there are neither marked nor consistent differences between mothers of schizophrenics and wives or other female relatives of schizophrenic males, or between mothers of schizophrenics and wives and mothers of males with other nonorganic disorders. These results are summarized as part of Table 1. Application of an extension of the median test (Siegel, 1956) indicates that there are no significant differences between mothers of schizophrenics and informants in the other four groups.

Since these findings were unexpected on the basis of the Freeman and Grayson research, an explanation was sought for the differing results from data available about the informants. Possessiveness was found to be negatively associated with education. These results are also summarized in Table 1. With the exception of the last item, relationships are significant at the .05 level. In particular, informants who have not completed

TABLE 1
SUMMARY OF CHI SQUARE VALUES FROM EXTENSION
OF MEDIAN TEST
($df = 4$)

Item	By Kin Role of In- formant and Diag- nosis of Pa- tient	By Edu- cation of In- formant
1. Parents should sacrifice everything for their children	3.78	29.29*
2. A child should feel a deep sense of obligation always to act in accord with the wishes of his parents	5.74	15.46*
3. Children who are gentlemanly or ladylike are preferable to those who are tomboys or "regular guys"	1.57	10.17*
4. It is better for children to play at home than to visit other children	7.26	3.78

* Significant at $<.05$ level.

grammar school tend to be possessive. Although the chi square value for the fourth item is not significant, even here twice as many informants without grammar school education "agree" or "strongly agree" with the statement in comparison with those who have completed grammar school. Negative correlations, although not as high, also occurred when possessiveness was related to rent and income, suggesting an association between any measure of socioeconomic status and possessiveness.

Freeman and Grayson do not report whether or not they found education associated with possessiveness. A correlation between these two variables is consistent with studies of social class differences in child rearing, at least with respect to the differing verbal expressions of lower and middle class respondents in such studies (Davis & Havighurst, 1953; Green, 1955). Freeman and Grayson (1955) indicate that "The mothers were an unselected group; they met only the criteria of visiting a schizophrenic son and being literate" (p. 46). In contrast, their control sample was obtained in the following manner: "The mothers were contacted indirectly by volunteer students (mostly nurses and aides) in several undergraduate psychology classes taught by one of the authors under the auspices of UCLA University Extension. The students were asked to distribute a questionnaire to a mother of their acquaintance..." (p. 46).

In view of the findings reported here (despite methodological differences between the studies) and the strong possibility of a sampling bias in the original investigation, socioeconomic status may well be the more cogent explanation of Freeman and Grayson's results. Clearly, further and

more careful research is required to support the clinical notion that attitudes of possessiveness on the part of mothers are a significant factor in the psychogenesis of schizophrenia.

SUMMARY

Four items from the Shoben Parent-Child Attitude Survey, measuring possessiveness, were included in a survey of female relatives of male mental patients. The results fail to support the work of Freeman and Grayson. A negative correlation occurred between possessiveness and education, and to a lesser extent, income and rent.

REFERENCES

- DAVIS, A., & HAVIGHURST, R. J. Social class and color differences in child rearing. In C. Kluckhohn, H. A. Murray, & D. M. Schneider (Eds.), *Personality in nature, society, and culture*. New York: Knopf, 1953. Pp. 308-320.

- DWORIN, J., & WYANT, O. Authoritarian patterns in the mothers of schizophrenics. *J. clin. Psychol.*, 1957, 13, 332-338.
- FREEMAN, H. E., & SIMMONS, O. G. Mental patients in the community: Family settings and performance levels. *Amer. soc. Rev.*, 1958, 23, 147-154.
- FREEMAN, H. E., & SIMMONS, O. G. Wives, mothers, and posthospital performance of mental patients. *Soc. Forces*, in press.
- FREEMAN, R. V., & GRAYSON, H. M. Maternal attitudes in schizophrenia. *J. abnorm. soc. Psychol.*, 1955, 50, 45-53.
- GREEN, A. W. The middle class male child and neurosis. In A. Rose (Ed.), *Mental health and mental disorder*. New York: Norton, 1955. Pp. 341-357.
- SHOBEN, E. J., JR. The assessment of parental attitudes in relation to child adjustment. *Genet. Psychol. Monogr.*, 1949, 39, 101-148.
- SIEGEL, S. *Nonparametric statistics*. New York: McGraw-Hill, 1956.

Received March 14, 1958.

A REDEFINITION OF SOCIAL DESIRABILITY

CLINTON B. DE SOTO, JAMES L. KUETHE,¹ AND JOHN J. BOSLEY

Johns Hopkins University

EDWARDS (1957) has recently shown that people agree well in judging items from personality inventories on a dimension that he terms "social desirability." The purpose of this study is to suggest that this dimension might better be called "desirability for well-being."

Edwards defines the dimension operationally in terms of the instructions he gives Ss who make judgments on it. There can be no quarrel with a strict operational definition. However, it appears that "social desirability" has more meaning than this to most people. For example, Fordyce (1956) gives as a preliminary definition, "Consensus judgments as to what behavior, feelings, and attitudes win social approval in American society" (p. 171). For Fordyce, clearly, the social desirability or undesirability of a trait refers to the social approval or disapproval it earns the person who possesses it. The instructions Edwards (1957) gives Ss who are to judge the social desirability of items also reveal this intent, in the examples of socially desirable and undesirable traits given the Ss and in the inclusion of the statement, "Remember that you are to judge the traits in terms of whether you consider them desirable or undesirable in others" (p. 4).

It is contended here that the study of social

desirability in the above sense is subject to the following difficulties:

1. Desirability for personal well-being is a far more salient property of most items on personality inventories than is social desirability.

2. The instructions given to Ss who are to judge social desirability have not adequately separated social desirability from other kinds of desirability, so that judgments which are supposed to be judgments of social desirability are often judgments of the more salient desirability for personal well-being.

METHOD

Edwards (1957) has developed a Social Desirability (SD) scale for which he says all the items have either obviously socially desirable or obviously socially undesirable scale values (p. 45). This scale consists of MMPI items on which 10 judges agreed perfectly as to their social desirability or undesirability (whether they should be endorsed or denied by someone giving a socially desirable self-description).

In the present investigation the 39 items from Edwards' short SD scale were presented to students from elementary psychology classes in the Johns Hopkins University for judgments of their desirability under three sets of instructions. In all cases, three judgment categories (desirable, undesirable,

¹ Now at Teachers College, Columbia University.

neither) were used, and the items were put in the third person, thus: "He is easily embarrassed."

One of the three sets of instructions was essentially Edwards' standard instructions (1957, p. 4).

Another set of instructions, intended to make clearer than do Edwards' instructions that the judgments are to express approval or disapproval, was phrased as follows: (This set will be called the "social approval" instructions.)

If you tend to like or admire a person *more* when you know he has a trait, circle the word "desirable" opposite the number for that trait.

If you tend to like or admire a person *less* when you know he has a trait, circle the word "undesirable" opposite the number for that trait.

If you do not tend to like or admire a person any more or any less when you know he has a trait, circle the word "neither" opposite the number for that trait.

A third set of instructions asked for judgments of the desirability of the traits for well-being, without reference to social approval or social desirability. These instructions, called the "well-being" instructions, went as follows:

If you tend to think a person is *more* well off when you know he has a trait, circle the word "desirable" opposite the number for that trait.

If you tend to think a person is *less* well off when you know he has a trait, circle the word "undesirable" opposite the number for that trait.

If you do not tend to think a person is any more or any less well off when you know he has a trait, circle the word "neither" opposite the number for that trait.

The SD items were judged under Edwards' instructions by 32 Ss, under the "social approval" instructions by 42 Ss, and under the "well-being" instructions by 41 Ss.

RESULTS AND DISCUSSION

Ss given Edwards' instructions judged items as desirable or undesirable in agreement with Edwards' judges, that is, in agreement with the SD keying, on 70% of their judgments on the average. Ss given the "social approval" instructions agreed with Edwards' judges only 54% of the time. The difference is significant at the .001 level by *t* test. Ss given the "well-being" instructions agreed with Edwards' judges 75% of the time, a figure not significantly different from that obtained with Edwards' instructions.

Majority agreement of Ss with the SD keying was obtained for 35 of the 39 items under "well-being" instructions, for 30 items under Edwards' instructions, and for 23 items under "social approval" instructions.

In general, then, Edwards' instructions seem to be in effect very much like the "well-being" in-

structions and rather different from the "social approval" instructions. And, presumably as a consequence of this, the SD scale seems to be made up of items on which Ss agree pretty well as to their desirability for personal well-being, but much less as to their social desirability.

A few items were almost uniformly judged desirable or undesirable under all sets of instructions. For example, "Most any time he would rather sit and daydream than to do anything else" was judged undesirable by 88% of the Ss under Edwards' instructions, by 86% of the Ss under social approval instructions, and by 100% of the Ss under well-being instructions. "He shrinks from facing a crisis or difficulty" was judged undesirable by 91% of the Ss under Edwards' instructions, by 90% of the Ss under "social approval" instructions, and by 95% of the Ss under "well-being" instructions. Sloth and cowardice incur social disapproval in addition to being judged undesirable for well-being. They seem to be genuinely socially undesirable.

In contrast, "His sleep is fitful and disturbed" was judged undesirable by 94% of the Ss under Edwards' instructions, by 88% of the Ss under "well-being" instructions, but by only 24% of the Ss under "social approval" instructions, while for "He is easily embarrassed" these figures were 72%, 85%, and 36%. Insomnia and ease of embarrassment, although clearly personally undesirable, do not seem to be genuinely socially undesirable.

Why should Edwards' instructions have been in effect like the "well-being" instructions, rather than like the "social approval" instructions, pulling out MMPI items on the wrong basis? Why should the slight ambiguity in his instructions have had this result?

The probable explanation is that items on personality inventories, including the particular MMPI scales from which the SD items were drawn, do *not* vary greatly in social desirability, but do often have clear and salient implications that the respondent does or does not have ailments and troubles of various kinds. People seem to be very perceptive of such implications of inventory items for health and well-being. When they are asked to judge the desirability and undesirability of items on a somewhat ambiguously defined dimension, these salient implications of many of the items determine their judgments, albeit inappropriately.

It is obvious that in some situations, as when people apply for jobs, good health becomes *situationally* strongly social desirable; it wins approval. And people are commonly observed to fake well-being in such situations.

In other situations, as in counseling and class-

room testing, effort to win social approval does not dictate this or any other manner of faking, since social desirability in the abstract is so inconspicuous and uncertain a property of the items. Thus, Edwards (1957, p. 22) found that whether or not Ss are required to sign their names to the Interpersonal Check List in a classroom situation has no important effect on their scores.

It would seem that scores in such situations are limited by what degree of troubles Ss can admit to themselves rather than by their effort to win social approval, a distinction akin to that between repression and suppression. This conclusion is in keeping with the finding by Eriksen and

Davids (1955) that individual differences in tendency to repress are related to scores on the *Pt* scale, which is highly correlated with the SD scale.

REFERENCES

- EDWARDS, A. L. *The social desirability variable in personality assessment and research*. New York: Dryden, 1957.
- ERIKSEN, C. W., & DAVIDS, A. The meaning and clinical validity of the Taylor Anxiety Scale and the Hysteria-Psychasthenia Scales from the MMPI. *J. abnorm. soc. Psychol.*, 1955, **50**, 135-137.
- FORDYCE, W. E. Social desirability in the MMPI. *J. consult. Psychol.*, 1956, **20**, 171-175.

Received November 21, 1958. Prior Publication.

FAMILY WORLDS

A Psychosocial Approach to Family Life
Robert D. Hess and Gerald Handel

How does a family function? How does a family make a distinctive life of its own while living according to the values of society? What is a "normal" family?

This book offers a new theory of the nature of family life, and illustrates this theory in five psychosocial portraits of different families. One family is fleeing the insecurities of the past, and has founded a protective world. The Lansons strive above all for harmony and equanimity, but pay a price. Another family is beset by emotional tensions, and each member builds a separate world for himself. The Newbolds are energetic and self-sufficient, excelling in leadership and extroversion. The last family has managed to create a closeness of emotional warmth in an atmosphere of cohesion and companionship.

Nearly all previous research inquiries into the family have been statistical studies of common patterns or clinical studies of grosser abnormalities. Here is a new framework for family study which applies established concepts and clinical procedures to families who come from a non-clinical setting. The authors have faithfully captured the interplay between family, personality, and social forces as each group develops its special way of coping with its problems.

308 pages

\$5.00



At your bookseller, or directly from the publisher:

The University of Chicago Press, 5750 Ellis Avenue, Chicago 37, Illinois
In Canada: The University of Toronto Press, Toronto, Ontario

THE NATURE OF HYPNOSIS: ARTIFACT AND ESSENCE¹

MARTIN T. ORNE

Harvard University and Massachusetts Mental Health Center

THE most meaningful present-day theories of hypnosis interpret hypnotic phenomena along three major lines: (a) desire on the part of the subject to play the role of a hypnotized subject (Sarbin, 1950; White, 1941), (b) increase in suggestibility (Hull, 1933), and (c) a further less well-defined category that is called by White "an altered state of consciousness" and by others, "cortical inhibition" (Pavlov, 1923), dissociation (Weitzenhoffer, 1953), etc. depending on their theoretical orientations.

The heuristic model of hypnosis that underlies this paper incorporates these three aspects. One of the hypotheses of the paper holds that much hypnotic behavior results from the subject's conception of the role of the hypnotic subject as determined by past experience and learning, and by explicit and implicit cues provided by the hypnotist and the situation. These varied role conceptions appear to be the source of most if not all of the inconstant patterns of behavior seen in the hypnotic state.

An increase in suggestibility may be viewed as an increase in motivation to conform to the wishes of the hypnotist. A second basic hypothesis to be tested thus proposes that, although increased motivation may be a constant accompaniment of the trance state, such increased motivation is by no means a

¹ This investigation was supported in part by a post-doctoral research fellowship from the Public Health Service.

I would like to thank Robert W. White for his encouragement and guidance and Milton Greenblatt for his support and his cooperation in the use of the facilities of the Massachusetts Mental Health Center. I am greatly indebted to Donald O'Connell, for his help in the pilot studies, and Ronald Shor and Theodore X. Barber for their invaluable assistance in running the first two experiments reported here. In addition, thanks are due Peter D. Watson for his aid in preparing and editing the manuscript. Finally, I would like to express my appreciation to Abraham H. Maslow of Brandeis University, Philip Noguee of Boston University, William McGill of the Massachusetts Institute of Technology, E. G. Boring and Ray Hyman of Harvard University, and the respective departments of psychology for making their classes available to me.

phenomenon unique to hypnosis but can be seen to operate in other experimental and life situations with equal force.

By experimentally controlling these two elements, role-playing and increased motivation, it is possible to investigate their sufficiency for explaining all aspects of the trance state and the extent to which still other concepts, such as an altered state of consciousness, are required.

The third aspect of hypnosis, the altered state of consciousness, presents the greatest problem for investigation, yet it has been felt necessary to include the concept in all attempts to explain the phenomenon. This residual aspect, which remains after increased motivation and role-playing are accounted for, may be regarded as the "essence" of hypnosis, with reference to which increased motivation and role-playing appear as artifacts.

Three related experiments are presented. The first is devoted to the effects of "role-play artifact" on the manifestations of hypnosis commonly seen clinically. It demonstrates that much of the complex phenomenon which we call hypnosis may result from (a) the subject's preconceptions of what hypnosis is, (b) implicit cues by the hypnotist as to what he thinks it should be, and (c) the particular techniques of trance induction. The second experiment demonstrates an aspect of role-play artifact that is introduced by a concrete experimental situation. It investigates cues that an experimental design may give about the role the subject is expected to play and demonstrates that in some instances an experimental result may more reasonably be accounted for on this basis than by invoking "trance effects." The third experiment is concerned with the effect of "motivation artifact" upon performance. It examines the claims of increased physical capacity in hypnosis and tests the hypothesis that this may be accounted for by increased motivation.

Table 1 gives a schematic representation of the author's working model of the hypnotic state.

TABLE 1
SCHEMATIC REPRESENTATION OF A WORKING MODEL OF HYPNOSIS

Situation of trance induction	"Role-Play Artifact" (cognitive component)	+ "Increased Motivation Artifact" (conative component)	+ Essence of Trance
Creation of situation to maximize: 1. Desirability of entering trance 2. Expectation that trance can be achieved 3. Respect and trust for operator 4. Restriction of extraneous stimuli 5. Focusing of attention	1. Expectations of Ss a. preconceptions b. cues from trance induction 2. Cues from Experimenter a. explicit b. implicit 3. Cues from experimental situation	The sources of increased motivation are not defined They represent a major area of future inquiry Probably some aspects will prove to be a component of "essence"	Uncertain
All techniques have the further qualities of: 1. Concrete suggestions in vivid simple language 2. "Suggestions" utilizing the perception of subjective events as their basis 3. Suggestions of gradually increasing difficulty to insure successful responses 4. Praising (rewarding) explicitly or implicitly the subject's positive responses			

PRECONCEPTIONS OF HYPNOSIS AND
THEIR EFFECT ON TRANCE
MANIFESTATIONS

The states induced by Mesmer (Binet & Féré, 1888; Boring, 1950), Coué (1922, p. 83), Wells (1923), Schilder (1956), and others are all hypnosis, yet their descriptions of how hypnosis characteristically manifests itself are very different. The common characteristics of these varied states that bring them all under the heading of "hypnosis" would appear to include: posthypnotic amnesia, apparent inability to use a given motor system when a functional paralysis is suggested, various sensory illusions including positive and negative hallucinations of all sensory modalities, apparent memory disturbances or improvements as well as reported increased control over autonomic nervous system functions. Whether all of these phenomena are necessarily part of hypnotic behavior will be discussed below. In any event, hypnosis is evidently characterized by the ability of the subject (S) in this special state to experience changes that are not normally found in response to similar cues in everyday life.

What, then, determines the particular trance

manifestations that an S shows on entering hypnosis? In terms of the model presented here, the answer may lie in role-play artifact. From this viewpoint, Ss who enter trance are motivated to play the role of the hypnotized S, and the precise manifestations of this role depend upon their perception of what it entails. Behavior of the S in trance is then determined by the S's preconceptions about how a hypnotic S acts, and the cues, both explicit and implicit, as to the desired behavior which the hypnotist communicates in the process of trance induction.

To test this hypothesis that conceptions about hypnosis held prior to entering the hypnotic state affect an S's trance behavior, a pilot study and a main experiment were conducted in which volunteer Ss were given the erroneous prior impression that catalepsy of the dominant hand (with the other hand flaccid) is a typical feature of hypnosis. This behavioral item was chosen because it satisfied a number of criteria. It is sufficiently unusual to have been reported had it ever been observed as a spontaneous characteristic of hypnosis; it is easily recognizable so that judgments of its presence or absence are unequivocal; and it is sufficiently plausible as a charac-

teristic associated with hypnosis that it would be accepted as such by the *S* population.

Especial care was exercised to eliminate possible effects of the bias of the experimenter by making it impossible for him to influence the results. It is easy to suggest to an *S* by implicit cues that he manifest catalepsy as part of the hypnotic state. Perhaps catalepsy of one hand might also be suggested during induction of the trance. Selection of catalepsy of the *dominant* hand avoids this possibility, as the experimenter had no way of knowing whether the subject was right- or left-handed until he asked for this information after the data on catalepsy had been gathered.

Pilot Study

An introductory psychology class at the Massachusetts Institute of Technology was given a lecture on hypnosis. Prior to the lecture, and without the knowledge of the class, two students had been hypnotized and given the posthypnotic suggestion that upon entering the trance subsequently, they would manifest catalepsy of one hand, the dominant hand. One student was right-handed and one student was left-handed. The class was then given a 25-minute lecture on the nature of the hypnotic state, at which point volunteers were called for in order to demonstrate the phenomenon. Of the 11 students who volunteered, the two who had been previously hypnotized were selected in a fashion that appeared random. They were again placed in trance, in a manner that appeared to be the initial trance induction, and simple trance phenomena were demonstrated, including one-handed catalepsy. Attention was called to the fact that the right-handed student had catalepsy of the right hand, and the left-handed student had catalepsy of the left hand. Immediately following this procedure, three more students from the same group of volunteers, who had not been hypnotized previously, were placed in trance.

A class of psychology students at Harvard were subsequently given the same kind of a lecture and demonstration, following which four *Ss* were hypnotized and tested for one-handed catalepsy.

All three of the M.I.T. experimental *Ss* gave good trance results, and all showed catalepsy of the dominant hand. One *S* was left-

handed. Of the four Harvard students who were hypnotized immediately after observing three demonstration *Ss* with catalepsy of the dominant hand, three manifested catalepsy of the dominant hand and one, catalepsy of both hands. All *Ss* were right-handed.

Main Study

In order to make it impossible for the experimenter to communicate his desire that the *S* demonstrate unilateral catalepsy, the main study was performed in a rigorous "blind" fashion. In this instance matched classes were used, each of which had received a lecture and demonstration of hypnosis. In one class the hypnotic demonstration included catalepsy of the dominant hand, while in the other this was omitted. The *Ss* were then tested in small groups, with members of both groups mixed randomly. The experimenter thus had no way of knowing which subjects should manifest one-handed catalepsy.

Procedure

The procedure of the pilot experiment was repeated with members of the introductory psychology course at Boston University with the inclusion of the control group. Instead of asking for volunteers, three *Ss* were employed who were introduced to the class as having taken part in prior research. The same three *Ss* were used for both sections of the course, to which essentially identical lectures were given. The demonstrations differed only in that in one section the three *Ss* manifested unilateral catalepsy, while in the other section this was not demonstrated. No students from either class were hypnotized at that time. Volunteers were solicited and subsequently tested in such a way that the experimenter had no way of telling which lecture they had attended until after the completion of the experiment.² All but two *Ss* were tested by an experimenter who was not at the lectures.

Trance depth was rated by the experimenter and an observer. The degree of consensus was high and in no case was there more than a one point difference. In case of disagreement both ratings are recorded. The ratings are rough clinical estimates based on the phenomena

² One of these *Ss* was tested the evening of the lecture. The remaining *Ss* were tested approximately one month after the lecture.

TABLE 2
TRANCE BEHAVIOR IN THE EXPERIMENTAL AND
CONTROL GROUPS

Subject	Catalepsy		Trance Depth	Dominant Hand
	R.	L.		
Experimental group				
1. M.S.	+	0	4	R.
2. M.K.	+	+	4	R.
3. R.L.	+	0	4	R.
4. C.L.	0	0	2	R.
5. S.T.	0	0	3	R.
6. A.L.	0	+	3	L.
7. O.B.	+	0	3	R.
8. S.R.	+	+	5	R.
9. B.T.	+	0	4	R. ^{a, b}
Control group				
1. D.L.	0	0	4	R.
2. W.O.	0	0	4	R.
3. M.R.	+	+	3	R.
4. L.P.	+	+	3	L.
5. B.Z.	0	0	3	R.
6. L.V.	0	0	4-5	R.
7. M.O.	+	+	3	R.
8. A.T.	0	0	3	L. ^b
9. W.M.	0	0	1-2	R.

^a This *S* was tested the evening of the lecture when he appeared unannounced along with a group of *Ss* who had previously volunteered. The experimenter did not know which class the *S* had attended until after the experiment was over.

^b *Ss* tested by author. I was not aware of which class these *Ss* had attended, in fact, I did not know until subsequently that they had been at the lectures.

which could be elicited from the *Ss*. A rating of 1 indicated no response; 2 implied eye-closure and only partial hand levitation without a positive response to "challenge" suggestions, i.e., you cannot open your eyes, or you cannot bend your elbow; 3 referred to positive responses to all challenge suggestions but inability to achieve hallucinations or posthypnotic phenomena; 4 was used to denote those *Ss* who responded to suggested hallucinations, gave simple posthypnotic phenomena, but did not achieve a good posthypnotic amnesia; 5 referred to "somnambulists" who could achieve all hypnotic phenomena easily, including complete amnesia.

Results

Of the nine *Ss* in the experimental group, five showed catalepsy of the dominant hand. Two showed catalepsy of both hands, and two showed no catalepsy. None of the control group showed catalepsy of the dominant hand, but three out of the nine *Ss* showed catalepsy

of both hands. Table 2 gives a summary of the findings.

Discussion

The item of behavior that was used is known not to occur spontaneously; its occurrence is significant if it is found at all. The results of the pilot and main experiments may be regarded as confirming the hypothesis that trance behavior is affected by the individual's conceptions about hypnosis held prior to entering the hypnotic state.

It would not be expected that all *Ss* would show this behavior. No truly naive *S* population is available, and many of the *Ss* had observed hypnosis prior to the demonstration. Some *Ss* should therefore have sufficient prior information to have formed very strong conceptions unlikely to be altered by the relatively mild attempt to manipulate these ideas experimentally.

That three of the nine *Ss* in the control group spontaneously manifested catalepsy of both hands is readily understood in view of the repeated testing for catalepsy, which they apparently interpreted as a cue to manifest the behavior. None of the control *Ss*, it should be emphasized, manifested unilateral catalepsy, indicating that no such desire was communicated by the hypnotist to the *S*.

This study has demonstrated for a single behavior item that trance behavior is affected by individual preconceptions about hypnosis. The results can be extrapolated to account for the apparently fixed qualities, not stemming from cues given by the hypnotist, that are reported in practically all present-day descriptions of hypnosis.

Thanks to the media of mass communication, it is relatively easy for a particular view of hypnosis to have gained wide currency and thus be found as a part of the general knowledge in which the *Ss* share. Such novels as *Mario and the Magician* (Mann, 1931) and *Trilby* (DuMaurier, 1895) have had very wide audiences and are known indirectly to almost all members of our culture. Uncounted articles and features about hypnosis have been disseminated to all levels of society. The picture of hypnosis that emerges in all of these is that of a passive *S* in a sleeplike state who has amnesia for the events occurring in hypnosis, and responds only to the hypnotist's sugges-

tions. According to Dorcus, Brintmall, and Case (1941), 79% of the student sample that they studied accepted hypnosis as possible, 71% had discussed hypnosis with someone, 54% had read about it, and 29% had actually seen a hypnotic trance at one time during their lives.

In the context of group tests for "suggestibility," in order to screen Ss, the investigators asked 57 students in elementary psychology courses: "Have you observed any other demonstrations of hypnosis; if so, where and when?" and "What have you read about hypnosis?" Only 12 Ss denied both having read about hypnosis and having had any chance to see the phenomenon previously; 18 Ss had seen hypnosis demonstrated in some form, and 23 had somehow read about it.

In the context of the questionnaires used in the above studies, "having read about hypnosis" meant specific reading in the scientific sense. In questioning well over 200 student Ss about their knowledge of hypnosis, the author failed to find one who did not have a very clear-cut notion about the nature of hypnosis, and who could not define the trance in a fashion similar to that found in dictionaries. Furthermore, they had all read something about hypnosis and could recall having done so, once it was made clear that this included nonscientific sources. The normal S population thus knows the meaning of the word hypnosis prior to taking part in any study.

CUES IMPLICIT IN AN EXPERIMENTAL DESIGN

An S participating in an experiment is aware that his responses are being recorded for specific purposes—that there is a *raison d'être* for the experiment—and he frequently has some idea of what these purposes are. How this knowledge affects the S's behavior depends upon the motivational structure that he brings to the experimental situation. The participation of the college student volunteer in psychological studies is usually due, not to the relatively low monetary remuneration but, rather, to his interest in taking part in scientific research, which in turn is likely to be based, at least in part, on a desire to further "progress in science" by his participation. Since the experimenter is perceived as knowing what he is

doing, furthering "progress in science" may well be equated with "making the experiment work" or, in more sophisticated terms, having his individual performance support the hypothesis of the experiment. Thus, when the S is motivated to comply with the wishes of the experimenter, his responses are readily influenced by what he perceives to be the basic hypothesis of the experiment.

Typically, the experimenter's hypotheses are not stated explicitly to the S because of the very considerations just mentioned. But unstated hypotheses may be conveyed implicitly by the experimental procedure itself, through what will be called here the "demand characteristics of the experimental situation." It should be understood that a person may fail to perceive fairly clear demand characteristics either because of lack of past experience or because of an inability to generalize from it.

Demand characteristics thus conceived appear central to much psychological work. Experimental situations vary widely in the extent to which they convey the purpose and the hypothesis of the experimenter. If an S can describe a hypothesis being tested, of which he is supposedly unaware, the experimental arrangements have significant demand characteristics. The obvious way to test for their presence is to ask the S about his perception of the experiment and its purpose. Usually, however, Ss are reticent about revealing their notions about the purpose of the experiment.

It is reasonable to assume that the student S population has some sophistication in regard to the philosophy of experimentation. They are aware that if an S is not told the purpose of an experiment he ought to remain naive in regard to it, lest his knowledge influence his performance. At the same time they understand the necessity for an experimental S to be "honest" in his response to the experimental situation and to questions about it. For these reasons, Ss are motivated to avoid recognizing explicitly the purpose of an experiment even though it may be clearly communicated by its design. Thus, the response to the direct question "What do you think this is about?" tends to be "I don't know." The S's behavior may nevertheless clearly betray an implicit aware-

ness of the relevant factors, and he may even verbalize them after the experiment in a "bull session" with his friends. We deal, therefore, with "knowledge" not readily available to consciousness which must be elicited in a clinical fashion. As in the case of other such material, the boundaries of consciousness may be expected to vary with the situation. When, however, a clinical approach is used in an inquiry and the *S* is pressed, one may be amazed—or horrified—by the *S*'s ability to formulate one's hypotheses in a lucid and at times highly sophisticated fashion. Unfortunately, the so-called inquiry is usually a most casual procedure.

While the demand characteristics of experimental situations probably have wider significance than is generally recognized, they are particularly significant for hypnotic experiments. Hypnotic *S*s tend to be particularly cooperative, almost eager participants. Furthermore, one of the assumptions of the present research for which there is extensive observational support is that the hypnotic state as such increases the motivation of the *S* to comply with the wishes ("suggestions")—both explicit and implicit—of the experimenter. The extent to which compliance can take place depends upon the demand characteristics in the experimental situation. The usual problem of demand characteristics (difficult enough to control in other fields of psychology because of the unconscious cooperation between *S* and experimenter) is thus compounded in hypnotic research.

In order to investigate the influence of the demand characteristics of an experimental procedure, a recent study (Ashley, Harper, & Runyon, 1951) was repeated with minor variations to be described. This experiment attempts to demonstrate a further dimension of the Bruner-Goodman (1947) effect, which has been the center of major controversy in recent years. Bruner and Goodman's basic tenet was that the perceiver's values alter his perception. There is no question that the perceiver's *previous experiences* may affect perception. A dispute, however, centers about whether *values* as such are significant variables affecting perception.

In order to show "clearly and unequivocally that the perceiver can contribute to the organization of his perception in a structured

stimulus-situation," Ashley, Harper, and Runyon (1951) argue it would be necessary to have a special situation. They state: "The Bruner and Goodman type of experiment would do this *if* the rich group and the poor group were identical in every other respect—in terms of their experience with money, their life histories, their physiological conditions, in short, if the sole difference between the two groups was that only one group had the psychological organization . . . of rich people and the other group the psychological organization of poor people." They go on to say: "Actually for our problem, it is irrelevant whether the *S*s are economically as well as psychologically rich or poor, or whether they are only psychologically rich or poor. In either case, a difference in performance of the two groups would reflect a difference in the perception due to the psychological organization of the perceivers" (p. 565).

In order to obtain two groups identical in every respect but for their perception of their economic status, they used hypnosis. While the *S* was in trance, artificial life histories were induced—one rich and one poor—each followed by induced amnesia. In essence, then, they view the situation as if two identically matched groups were available—one rich, and one poor. It is assumed that because amnesia was induced for the preceding state, the *S* is again naive and that the only difference is in respect to his perceived economic status.

The final sentences of their rationale are particularly interesting. "Even though we do not know fully what happens when we hypnotize a person, if we do hypnotize him and tell him he is rich and he behaves in one way in the coin-matching situation, and then, a few moments later, we tell him he is now poor and he behaves in another way, *we can conclude that the observed difference is due to a change in his psychological organization*" (Ashley et al., 1951, p. 565).³ The authors in fact conclude from their data that the psychological organization (including the wants, needs, interests, attitudes, and values) of the person contributes to the figural organizations of his perceptions.

It is unquestionably true that observed differences in coin-size judgments are due to

³ Italics mine.

changes in psychological organization. The question with which we are concerned, however, is whether these changes in psychological organization relate to the actual experiencing of the feelings of being rich or poor, or whether they reflect the demand characteristics of the experimental procedure. The hypothesis to be tested is that the demand characteristics of the experiment are largely responsible for the results obtained by Ashley et al. (1951).

Disregarding the theoretical framework of the study, this is what actually took place: An individual was told—in hypnosis—that he was very poor, then—again with amnesia in hypnosis—that he was very rich and, subsequently, with another hypnotically induced amnesia, that he was himself. In each of these states he was required to make a series of coin-size judgments. The authors' interpretation rests largely on the assumption that hypnotic amnesia is truly the same as not knowing. Granted this, one would be justified in ignoring the fact that the procedure of coin-size estimation is repeated and that economic status is hypnotically induced. However, data are available that lead one to question this assumption.

One of the few specific experiments dealing with posthypnotic amnesia directly is a study by Strickler (1929), who compared the relearning of nonsense material in the posthypnotic state with induced amnesia with the learning time required for the material not previously learned. He concludes that "the posthypnotic amnesia ordinarily met with, which appears superficially to be a complete wiping-out of memory, is by no means complete."

Even more relevant are the data obtainable in hypnotic age-regression. Here we are dealing with an induced amnesia in hypnosis for what purports to be all material learned after a given age. All studies of hypnotic age-regression have shown that some material persists no matter how "real" the regression appears.

In the investigator's prior work (Orne, 1951), it was possible to show that an individual regressed to age six was able to comprehend English, though he himself pointed out in German that he could not understand it. Historically, the *S* was unable to understand English at age six. Another *S* could spell with-

out error "I am conducting an experiment which will assess my psychological capacities." Another was able to give the square root of four, and so on. Furthermore, if we test for amnesia in a more subtle fashion, it is easy to demonstrate in the wake state or in trance that no true ablation of the material for which the *S* has amnesia exists, despite his subjective feeling of being unable to remember.

The fallacy of the assumption that knowledge for which the *S* has amnesia does not influence his behavior can be seen in any posthypnotic suggestion. The *S* firmly denies recall yet assiduously persists in the suggested behavior. The phenomenon is well known in response to an explicit cue; it would seem rather absurd to deny it in response to an implicit one.

A pilot study was therefore conducted that replicated all essential characteristics of the Ashley, Harper, and Runyon experiment, with the addition, however, of a careful inquiry after the completion of the experiment. The procedure was patterned after the inquiries commonly performed as part of the Rorschach test, which seek answers to a series of questions without providing the *S* with a cue as to the answers expected. 1. The subject's perception of the experimental task was elicited by a general question, "What do you think this experiment was about?" 2. The *S*'s perception of the purpose of the investigation was elicited by questions such as "What do you think this experiment is trying to prove or demonstrate?" 3. *S*'s perception of the experimenter's hypothesis was elicited by direct questioning, with such questions as "What do you think I hope to find?" 4. The *S* was also asked about his own hypothesis concerning the study—what he, on the basis of what he knew about the experiment, would predict the results to be. 5. The final question related to his beliefs about his own performance with the question, "What do you think your experimental behavior demonstrates?"

The following hypotheses were formulated:

1. The subject in an experiment is usually able to express some demand characteristics of the procedure, if careful inquiry is conducted and his initial resistance is penetrated in a clinical fashion.

2. The majority of subjects may perceive the same demand characteristics in the experi-

ment and these may be the same as the hypothesis being tested.

3. These demand characteristics rather than the experimental variables may be the major determinant of the subject's behavior.

a. If the majority of subjects perceive the same demand characteristics, then subjects who fail to perceive them should not show the behavior characteristic of the group.

b. If the demand characteristics are the determinant of subjects' behavior, it is possible for an experimental design that omits a crucial aspect of the original independent variable to elicit similar responses to the extent that the same demand characteristics are present.

Pilot Study

The pilot study was designed to test the first two hypotheses.

Procedure

The Ashley, Harper, and Runyon study was repeated in all essential details with four undergraduate Ss, with the addition of appropriate inquiry. Equipment employed in the original Bruner-Goodman study (1947) was used for making the coin-size estimations. Unlike the procedure of Ashley, Harper, and Runyon, however, the coins were presented on the S's left palm which he was permitted to hold beside the box. He was not permitted to remove the coin from his palm.

All Ss used in this study had demonstrated their ability to manifest all of the usual deep trance phenomena including responsiveness to posthypnotic suggestions and the ability to experience what appeared to be total amnesia when this was suggested.

The procedure, briefly stated, was as follows: After the S was placed in trance, amnesia for his own life history was induced. He was then given a pseudo-life history which was essentially the same as that described by Ashley, Harper, and Runyon. The poor state was induced first, then the rich state, and finally the normal state. The S judged coin sizes in all three states. The same S was run with both imagined coins and with real coins presented in all three states. Also in all three states, he was given brass slugs which were called "lead," "silver," "gold," and "platinum." The brass

was of a very whitish color so that it could conceivably have been the appropriate metal.

Results

The results are summarized in Figure 1b, which presents the subjects' average coin-size estimates. The data are essentially identical to those obtained by Ashley, Harper, and Runyon (see Fig. 1a). The data on the size estimates of "slugs" successively called silver, gold, and platinum were also similar to those presented by Ashley, Harper, and Runyon in their series using a lead slug. All of the four subjects were able to describe correctly the purpose of the experiment and the hypotheses of the investigator who originally designed the experiment.

Discussion

The data from the pilot study imply that the present procedure effectively reproduces that of Ashley, Harper, and Runyon. Both in terms of the quantitative results and the observed behavior of our S, no significant differences emerge.

The only essential difference between these data and those obtained by Ashley, Harper, and Runyon relates to the inquiry procedure. The results confirm the first two hypotheses.

1. The S in an experiment is able to express some demand characteristics of the procedure, if careful inquiry is done and his initial resistance is penetrated in a clinical fashion.
2. The majority of Ss may perceive the same demand characteristics of the experiment and these may be the same as the hypothesis being tested. However, the third hypothesis has yet to be dealt with.

It is interesting to note that two of the four Ss who were specifically questioned about this point denied vehemently that they were influenced during the experiment by an awareness of the experimenter's hypothesis. But the S's verbalization during inquiry cannot be accepted at face value. As long as the S recognizes and is able to verbalize the demand characteristics of the experiment, they may play a significant role in his experimental behavior, although to demonstrate that they do so requires supporting evidence. It is with this further evidence that the main study is concerned.

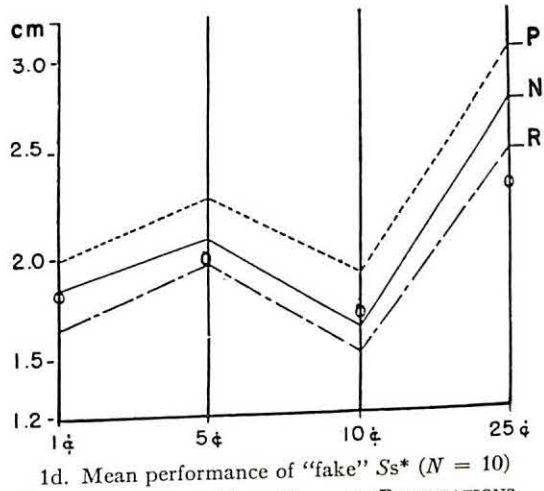
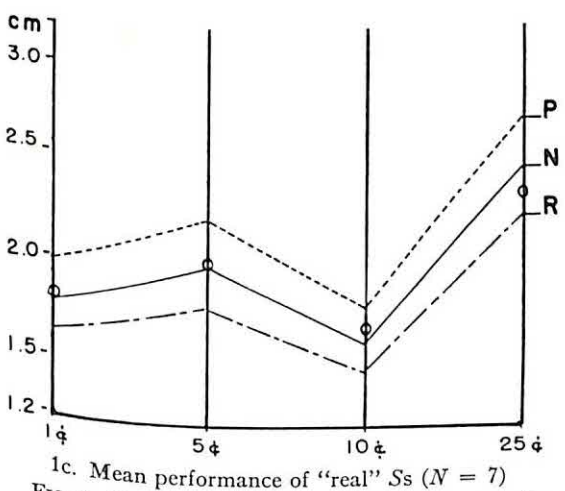
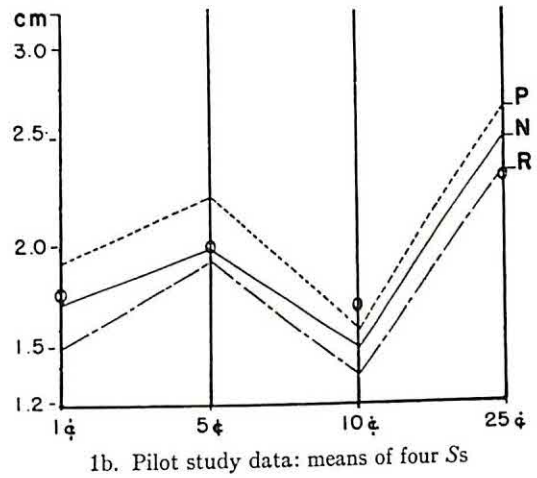
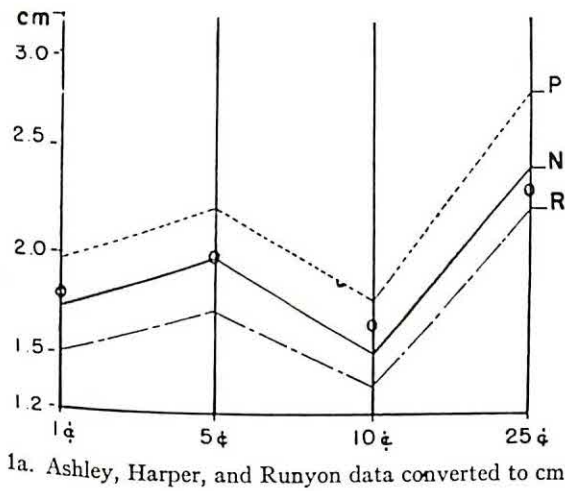


FIG. 1. COMPARISON OF ASHLEY, HARPER, AND RUNYON DATA WITH DATA FROM PRESENT REPLICATIONS
(Legend: [R] = Rich; [P] = Poor; [N] = Normal; [O] = Actual coin size)

* One S (B. S.) was highly atypical and therefore excluded. See Fig. 2(b) for his performance.

Main Experiment

While the data found in the pilot study are consistent with the hypothesis that the demand characteristics of the experimental procedure may determine behavior, they are open to several serious objections.

The greatest single problem relates to the technique of inquiry and the interpretation of the data obtained in this fashion. It is important to have an objective method of rating how well the S perceives the demand characteristics of the experimental situation. The study was therefore designed so that the S's inquiry would be rated by independent judges who did not have available to them the S's data, but who would only have the opportunity of reading transcripts of the inquiry.

Another problem is a bias inherent in the inquiry procedure. Some Ss who do not perceive the demand characteristics while engaged in the formal experimental procedure may perceive them during the inquiry. In such a case, and if the demand characteristics rather than the experimental variables determine the response, then the inquiry may indicate that the S should have responded a certain way when in fact he did not. However, the reverse should not occur.

The question still remains as to whether the S's perception of the demand characteristics is responsible for his behavior, or whether it is due to the operation of the "intended" experimental variables. This question was dealt with by including a control group

that could not conceivably be construed as experiencing a "psychologically rich and poor state." If it could be demonstrated that a group of *Ss* who do not experience the "rich and poor state" but are exposed to the demand characteristics of the procedure also show the data reported, it would be justifiable to attribute the results to the demand characteristics rather than to a presumed change in the psychological organization of the individual because of being "psychologically rich and poor." The control group thus permits inferences without reliance on the inquiry.

A group of *Ss* who were not in hypnotic trance and did not manifest amnesia should provide such a control group. They would, of course, have to go through the same procedure as the "real trance" group. Such a group of *Ss* would be asked to "play act" being in hypnosis and go through the whole procedure as if they were real *Ss*. This group of *Ss* would not truly consider themselves as psychologically rich or poor. In these *Ss* no amnesia could be induced, and their behavior would clearly be that of a group of persons acting under three different sets of instructions—act as though you were poor, rich, and yourself.

This type of procedure is open to an important objection. Experimenter bias could play a major role. While the procedure and the wording of instructions would be the same, it would be possible unwittingly to include a variety of cues which could differentially shape the behavior of the two classes of *Ss*. A blind technique is thus necessitated, in which the experimenter would not know which *Ss* were "real" and which were "fake."

Such a stratagem presupposes that a "fake" *S* can simulate hypnosis sufficiently well to deceive the experimenter. However, there is a widely held opinion in the literature that it is impossible to simulate hypnosis successfully (Jenness, 1944; Stokvis, 1955). Cursory attempts by the author to have *Ss* fake trance showed that the *S*'s efforts were half-hearted and obviously transparent.

In the usual faking situation, the experimenter knows that the *S* is faking, and the *S* is aware that the experimenter knows it; the usual purpose of this situation is to demonstrate the difficulties of fooling an experienced hypnotist. Clearly, the experimenter is not

really expecting the *S* to be able to carry out his task effectively, and the *S* is aware of this. Thus, the *S*, who is anxious to please the experimenter, is in actuality motivated to give an unsuccessful performance. Furthermore, since the *S* is aware that the experimenter knows that he is acting, the *S* feels, with good cause, that it is impossible to deceive the hypnotist. There is a marked tendency to smile during induction procedure and in response to suggestions that might be construed as foolish, as well as to ask "How am I doing?" at intervals. Any suggestions that evoke even mild discomfort are followed only briefly and half-heartedly.

Most classical texts and modern authorities agree that hypnosis cannot be faked easily and "if a subject attempts to fake, tests for anaesthesia will permit ready recognition" (Estabrooks, 1948; LeCron & Bordeaux, 1947, p. 103; Mayer, 1951). However, the author has, upon two occasions, been taken in by *Ss* who had apparently faked their way through the procedure and who subsequently disclosed the fact. In discussions with other hypnotists, he found that all who had had considerable experience could recall similar instances.⁴ These experiences are usually explained by stating that such *Ss* must really have been in the trance state or they would not have been able to act as well as they had, a view that is supported by the literature on hypnosis. The report of the *S* that he has not really been in hypnosis is thus lightly passed over, since "subjective reports are really not reliable." It is the author's opinion that it is dangerous to ignore the conviction of the *S*, expressed in good faith, that he did not experience the amnesia or anaesthesia or any other trance phenomena that he appeared to have experienced; and that it is indeed possible to construct a situation that would facilitate the successful "faking" of a hypnotic trance.

To do so it is necessary to motivate the *S* to

⁴ In this respect, it is interesting to note that Pattie (1937) was sufficiently concerned with being deceived by a faking *S* that he asked each *S* to sign a statement "I, realizing that the experiment performed on me will probably be published in a scientific journal, solemnly declare that I was not faking or imitating the hypnotic trance but that I was genuinely hypnotized and do not remember the events of the experimental periods."

act the part of a hypnotic *S*—not to go into a trance—and for him to believe that it is possible to convince the hypnotist that he really is in trance. Accordingly, a situation was set up in which the *S* knew that the hypnotist would not know in advance whether or not he was really in trance. Further, it was so structured that the *S* would perceive the desire of the experimenters for him to do a successful job of faking. Finally, the situation was presented as a challenge that provided ample motivation for the *S* to tolerate unpleasant suggestions or tests.⁵

Training of Subjects

College students from Harvard and Brandeis were used as *Ss*. All *Ss* had been present during a lecture about hypnosis and a demonstration of the phenomenon given by the author in the context of an introductory psychology course. All *Ss* were volunteers who agreed to take part in the study at the time of the lectures. No *Ss* were paid. All had also been tested and rated by investigators other than the author for "suggestibility," in at least one screening session. All *Ss* in the real trance group had further training with the author's associates. In order to satisfy criteria for depth of trance, they had to manifest the ability to respond to all the usual hypnotic suggestions with special emphasis on analgesia, visual hallucinations, negative auditory hallucinations, amnesia induced by suggestion, and simple posthypnotic suggestions.

The fake group of *Ss* was drawn from the same population as the others but did not have special training in entering trance beyond the first screening session. They were all informed of their task in the manner described below.

The experimental procedure was administered to all the *Ss* by the author, and he was

given no information or cues by his associates as to which group any *S* might belong. Nor was he informed of the number of real and of fake *Ss*. There were seven real *Ss* and 11 fake *Ss*.

Special Procedure for Faking Subjects

Prior to the experiment, the faking *Ss* were told by an investigator other than the hypnotist:

We're doing a study in which we would like you to take part. You recall some time ago you volunteered as a subject for some hypnotic experiments. We would like you not to go into a hypnotic trance; however, Dr. Orne, who will be working with you, will not know whether you are a real or a fake subject. He will go through trance induction with you, and then several tests and an experiment. We would like you to act as if you were entering a trance, but I don't want you to go into a trance. Your job is to fake the whole thing and to try to fool Dr. Orne. He knows that I may have some people trying to fake their way through the experiment but he has no idea who they are. I would like you, as a special favor, to try to do the best possible job of fooling him; to really convince him that you are in trance. This is quite difficult but a very important part of the experiment. I am trying to find out whether people are capable of doing this successfully. One of the tests which he will do with you will involve a harmless but somewhat painful electric shock. After the trance part, which you will have been faking, he will ask you some questions about it. Again, I don't want you to let on that you have been faking, and I want you to try to answer these questions as though you had been in hypnosis. Please keep on playing this part until I give you a signal. After the signal you can tell him what really happened. The signal will be that I will turn off the light on a box which will be on the table in front of you. I am sure that you will be able to do a good job of it.

The *Ss* usually objected by saying that they did not know how a hypnotized *S* would act. They were told that although this feeling was understandable, others had been able to carry it off and that besides, they themselves had seen other *Ss* go into a hypnotic trance during a previous demonstration.

The motivation for the *S* to act the role of the hypnotized individual is fairly clear. The author had been presented to the *Ss* as an authority in the field of hypnosis and they had been taken into confidence by one of his associates in an attempt to fool him. Thus, on the one hand they were contributing to the advance of knowledge, on the other, they would have the pleasurable experience of "putting one over" on one of their teachers.

⁵ After the completion of this study a previous reference to the use of a faking subject as a form of control was encountered. In an experiment on hypnotic pain suppression, Dynes (1932) asked a member of the psychology department to talk to *Ss* privately and have them fake a trance the next time. The *Ss* thought that he knew nothing of the plan and carried out the request to fake. However, Dynes himself did know and reports no difficulty in distinguishing this state from the real trance. It would have been interesting to know whether Dynes would have found it quite so easy to distinguish the behavior had he not known in advance.

Experimental Procedure

At the experimental session the *S* was exposed to trance induction. The Erickson hand levitation technique as described by Wolberg (1948) was used. After the *S* appeared to be in trance, a series of tests were done such as: hand clasp, analgesia, negative auditory hallucination, and positive visual hallucination.

When these tests had been completed, the *S* was told that his name would appear on his forehead and that he would feel it as though it were there in Braille. After this suggestion had been accepted, he was informed that if he would wipe this off he would forget everything about himself. Once he had accepted the suggestion of total amnesia, the rich state and the poor state was induced in a fashion identical to that used by Ashley et al. (1951). The *S* was then asked to make coin-size judgments of a penny, nickel, dime, and quarter, again using the original Bruner-Goodman box. He was required to make six coin-size judgments of each coin in each state. The final set of judgments was made in his normal state, but with the suggestion that he had no recollection of anything that had happened since he entered trance. The rich and poor states were randomized as well as the order of the coins. A balanced design could not be achieved in respect to coin order because of the number of *Ss* used, but it was approximated as far as feasible. The hypnotist-experimenter did not know the key to the order and was merely told in each instance which state he was to run first and with which coin to start.

The procedure, used by Ashley, Harper, and Runyon and in the pilot study, of estimating the size of the lead slug which was called lead, silver, gold, and platinum, was omitted.

At the completion of the coin-size judgments, an inquiry was performed as previously described. This inquiry was the same for all *Ss*. At the completion of the inquiry, both in the hypnotic and wake states, the light of the Bruner-Goodman apparatus was turned off, which was the signal for the fake group to tell what really happened. If the *Ss* were real, this had no meaning to them, and the experiment was completed. Those of the fake group were asked a series of questions as to whether or not they had been in trance. They were also asked about their performance, why they had acted as they did on some points, and if and

when they felt anxious. Finally, they were asked whether they thought that they had given themselves away; if so, when and why. All of these inquiries were completely tape recorded and transcribed so that, subsequently, judges could rate what each subject perceived as the hypothesis being tested.

Results

1. *Objective Findings-Group Results.* Figure 1 displays a comparison between the present data and those of Ashley, Harper, and Runyon (Fig. 1a); in Fig. 1c, the results of the hypnotic *Ss* are averaged in the manner described by Ashley, Harper, and Runyon; in Fig. 1d, the results of the "fake" *Ss* are averaged in the same manner. All three graphs present essentially the same configuration. In all cases the judgments in the poor state are the largest, judgments in the rich state smallest, while judgments in the "normal" state fall between.

2. *Results for Individual Subjects.* Figures 2a and 2b give the results for each "real" or "fake" *S*. *Ss* varied widely in their response to the experimental task, not all of them yielding a configuration that corresponds to the group average. While Ashley, Harper, and Runyon unfortunately do not give their individual results, they report considerable variation.

3. *Comparison of Judges' Ratings with Objective Categorization.* Using analysis of variance for each individual *S*, it is possible to test statistically whether there are significant differences between *S*'s coin-size estimates in any combination of the three states and the direction of significant differences. Ignoring the "normal" judgments, the possibilities reduce to three categories: no significant differences between rich and poor, poor significantly larger than rich, and rich significantly larger than poor. Each *S*'s coin-size judgments were classified into one of the three categories on the basis of statistical analysis considering differences not significant at the .05 level as no difference.

The transcribed postexperimental inquiries were given to two independent judges to rate the *S*'s perception of the hypothesis being tested at the time of the experiment in terms of the same three categories. The judges had no contact with the *Ss* or each other. Table 3 shows a comparison in terms of the

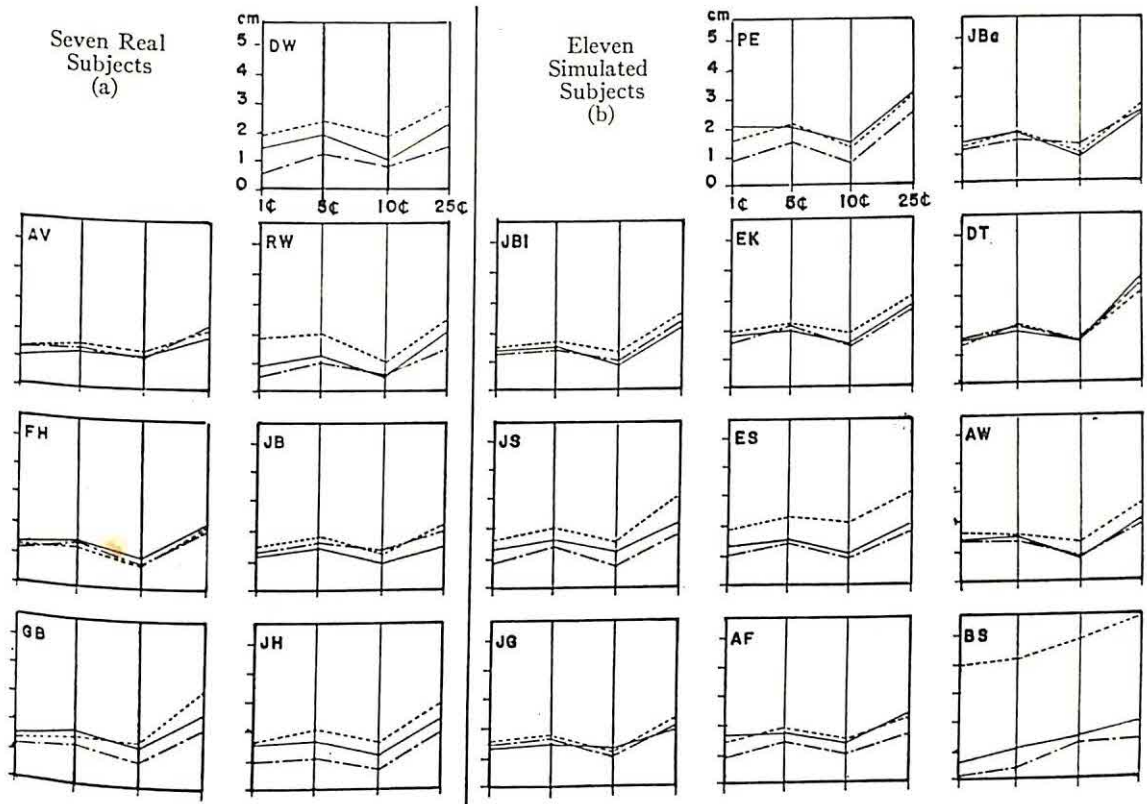


FIG. 2. GRAPHS OF INDIVIDUAL SUBJECTS

(Legend: [----] = Poor; [—] = Normal; [— · —] = Rich)

TABLE 3
A COMPARISON OF JUDGED AND ACTUAL CATEGORIES
OF RESPONSE

Subject	Major Categories		
	J1	J2	Actual
G. B.	II	II	II
F. H.	II	II	I
J. H.	II	II	II
A. U.	II ^a	I	I
R. W.	II	II	II
D. W.	II	II	II
J. Ba.	I	I	I
P. E.	II	II	II
J. Bi.	II	II	II
J. G.	I	I	I
A. W.	II	II	II
E. S.	II	II	II
J. S.	II	II	II
D. T.	I	II	I
A. F.	II	II	II
E. K.	II	II	II
B. S.	II	II	II

NOTE.—Key to symbols: I. No significant difference; II. Poor significantly larger than rich; III. Rich significantly larger than poor.

Two judges were used (J1 and J2). Note that no S actually belongs in Category III and that neither judge placed any S within it.

^a Judge undecided about I or II here but chose II as better estimate.

three categories between the ratings of the two judges and the individual's responses. There is a high degree of correspondence between the judges' ratings and S's performance.

Testing the null hypothesis of no systematic correspondence between judges' rating and S's coin-size judgments leads to its rejection for each judge (.015, Fisher Exact Test); combining the significance levels of the two judges leads to an over-all significance of less than .01.

Discussion

The data obtained from the seven hypnotized Ss are essentially identical to the findings of Ashley, Harper, and Runyon, and virtually indistinguishable from the performance of the 11 stimulating Ss. These results confirm the hypothesis that it is possible for an experimental design that omits a crucial aspect (hypnotic amnesia) of the original dependent variable to elicit similar responses as long as the same demand characteristics are present.

The subjective experience of members of the stimulating group was radically different from that of the Ss in deep trance. The Ss readily

described their conscious efforts to "second guess" what the experimenter would expect of them if they were actually in hypnosis. The data obtained from the simulating group are the result of a concerted effort on the part of the Ss to respond in a way identical to hypnotized Ss. The subjective experience of the hypnotized group was different. While clinical inquiry revealed the Ss' perception of the author's expectations, they denied that these factors had any effect upon their performance. This denial on the part of the hypnotized S does not, of course, mean that their perception of the experimental purpose was unimportant. It does mean, however, that they themselves were not aware of its significance.

An investigation of the demand characteristics perceived by each S may account for individual results that did not conform to the group average, as an examination of the judges' ratings confirms. It was discovered that the inquiry procedure had not been refined sufficiently to permit prediction of the Ss' performance in the "normal" state. However, performance in the "rich" and "poor" states could be predicted with a high degree of accuracy from the judges' ratings of the Ss' perception of the experimental purpose. No S reversed the expected trend by making his coin-size judgments larger in the rich state than in the poor state. No S was rated by either judge as having perceived this to be the hypothesis of the experiment. Twelve subjects made the coin-size estimates significantly larger in the poor state than in the rich state. All 12 Ss were rated by both judges as having perceived this to be the author's hypothesis. Five Ss failed to significantly differentiate their coin-size judgments between the rich and poor state. Of these five subjects, four were rated by either one or both judges as having failed to perceive the demand characteristics of the experiment.

The inquiry data thus support Hypothesis 3a, that if the majority of Ss perceive the same demand characteristics, then Ss who fail to perceive these demand characteristics should not show the behavior characteristic of the group.

The present experiments do not bear on the validity of the Bruner-Goodman effect. The Ashley, Harper, and Runyon experiment was

used, rather, as an example of a study that appears methodologically sound, but in which demand characteristics seem to be the major determinant of the S's performance. The implications seem clear: demand characteristics may determine behavior in hypnotic experiments. Before an effect can legitimately be attributed to hypnosis, it is necessary to demonstrate that it is not primarily a function of demand characteristics. Such proof appears to require the use of blind techniques and adequate inquiry.

THE INFLUENCE OF MOTIVATION ON HYPNOTIC BEHAVIOR⁶

In studying the nature of the hypnotic trance, the question arises as to which phenomena are primary and consistent components of the trance state and which are secondary derivatives. Let us postulate that increased motivation is a constant accompaniment of the hypnotic state. The present phase of the research was designed to show that certain phenomena long viewed as part and parcel of the hypnotic state may more parsimoniously be viewed as derivatives of increased motivation, and can be reproduced *pari passu* by other motivational techniques that have no direct relationship to hypnosis.

For years it has been claimed that there is an increase in physical capacity during the trance state. In part this claim has been based on casual observation, the favorite example being that of the stage hypnotist who places a subject in deep trance across two chairs and permits one or more individuals to stand or sit upon him. This "experiment," with variations, is often cited as irrefutable evidence for increased physical capacity. Another group of frequently cited observations are those concerning the ability of the subject to maintain his hand in an outstretched position for extended periods of time without evidence of fatigue. On the basis of this type of data, estimates of greatly increased physical capacity have been made (McDougall, 1926; Moll, 1904).

An early study by Nicholson (1920, p. 89) maintained that "during the hypnotic sleep the capacity for work seemed practically end-

⁶ This experiment was originally reported in German (Orne, 1954).

less." Unfortunately, no quantitative data were given, and the study was poorly controlled. In a meticulous investigation, Williams (1930) showed no difference between hypnotic and wake states in the ability to maintain the arm in an outstretched position. However, this study failed to employ suggestions to the effect that the arm would not get tired and could not drop. In another similar investigation, using an ergograph and employing appropriate hypnotic suggestions, Williams (1929) found a 12 to 16% increment in the trance. More recently, Roush (1951) showed an increment in performance in hypnosis significant at better than the .05% level using the arm dynamometer, the hand dynamometer, and hanging by the hands, as measures of fatigue.

All the experiments performed by psychologists in the laboratory have followed orthodox scientific methods insofar as a standard set of instructions was given to the *S* to hold a weight, pull an ergograph, or perform a similar task in both the nonhypnotic and hypnotic states. The better experiments used the usual ABBA arrangement to control fatigue or practice effects. Any increment in performance was defined as an increase in capacity due to trance. It is necessary here to question the logic on which the interpretation of these results is based. While these experiments undoubtedly show that instructions given in trance state result in increased performance over that achieved by the same instructions in the wake state, they do not necessarily show an increase in capacity. Alternatively, the *S* may be more willing to exert himself while in hypnosis. The governing factor could be the increase in the *S*'s motivation to comply with the experimenter's request rather than an increased capacity to comply. The instructions, while identical in wording, may be experienced as quite different by the *S* in hypnosis and the waking state. The request to hold a weight at arm's length, given in trance, may be a highly motivating cue or "suggestion," especially if the *S* is told that he is to feel very powerful and not fatigued. The identically worded request in the wake state is perceived as a request by the experimenter and may be followed if good rapport exists between experimenter and *S*. However, as the discomfort of the task in-

creases, the *S* becomes increasingly disinclined to comply. Viewed in this context, the reported experimental results do not necessarily imply that physical capacity is in fact increased in trance but, rather, that the trance state increases performance.

Procedure

Nine *Ss* in deep trance were asked to hold a kilogram weight at arm's length. This was done in such a way as to derive maximal benefit from the peculiar nature of the trance state. Thus the *S* was told to hallucinate a table, and only after the table was both seen and felt by the *S* was the suggestion given that the right arm would feel no fatigue and no pain.

All the standard tests of deep trance were met in each *S*. A kilogram weight was placed in the *S*'s right hand, and the *S* was instructed to place it on the imagined table, to continue holding it with his fingers, and under no circumstances to drop it or his arm. Continuous suggestions were given to the effect that he would be able to hold onto the weight, that his arm would not get tired, etc., and that he would continue to see the table. The end point was when the *S* was no longer able to hold up his arm and began to come out of trance. At that point he was reassured, told to drop the weight, and deep trance suggestions were again given. After some minutes, and having made certain deep trance was again established, the *S* was awakened with a carefully induced posthypnotic amnesia. The *S* was not told the length of his performance.⁷ For the second part of the experiment, which was done within half an hour of the first, the *S*, not now under hypnosis, was instructed as follows:

This is a most important part of our experiment. It is very important for us to know your endurance and

⁷ In the preceding section it was pointed out that the posthypnotic amnesia induced in hypnosis is not tantamount to an ablation of memory. One may be justified in assuming that the *Ss* do not know their hypnotic performance, not because of the amnesia but, rather, because they were never informed of the length of time they held the weight in hypnosis. A common belief that the *S* in hypnosis has a perfect sense of time would lead to the conclusion that this is not an adequate safeguard. Fortunately, a very thorough study of the time sense under hypnosis was conducted by Guenther Klaus in a doctoral dissertation (University of Freiburg, Germany, 1948) which demonstrates unequivocally that the time sense is not improved by hypnosis.

TABLE 4
COMPARISON OF SUBJECTS' PERFORMANCE IN HYPNOTIC
AND WAKE STATE

Subject	Hypnosis		Waking	
	Minutes	Seconds	Minutes	Seconds
1	4	05	5	33
2 ^a	4	40	6	25
3	4	38	8	06
4 (a) ^b	6	05	3	29
(b)	5	50	10	02
5	7	07	7	57
6	10	05	16	00
7	4	52	5	49
8	5	20	5	32
9 ^c	4	57	2 (a)	10
			5 (b)	09

^a This experiment was performed in 1950. In 1957, it came to my attention that this S feels that he simulated completely throughout this experiment. At the time, I was totally unaware of this possibility and the S was in trance by all the usual criteria.

^b S dropped the weight after 3' 29" in the wake state. The next day, care was taken to motivate him adequately. While the hypnotic performance was only 15" below the previous day, his wake performance now exceeded 10'.

^c This S suddenly dropped the weight without warning in the wake state after 2' 10". She was encouraged and after a 20' time lag again held the weight. This time her performance was 5' 09". This performance in itself is better than her hypnotic performance of 4' 57"; however, it might seem that the waking performance was better than this, as the 2' 10" period was not given credit.

physical capacity. What I want you to do is a very difficult task. It does not look difficult but it is. I want you to hold this kilogram weight at arm's length. Your hand will get tired and it will take great effort to do this. There is a natural tendency to drop the weight if your hand gets tired. However, it is vital that we get your true capacity. Surprisingly enough, our female subjects have been able to hold the weight for T minutes. [The time T given would be his previous performance during hypnosis rounded off to the nearest half minute.] Our male subjects have been able to hold the weight at least T + ½ minutes. I realize that this is a difficult and painful task. Just to make it interesting we will try a game. At T minus 2 minutes we will start you off at 5 cents. At T minus one and a half, we will double that and make it 10 cents. At T minus one, 20 cents. At T minus one half, 40 cents. At T, 80 cents and at T plus one half, \$1.60.

Then the S was told that while we could not afford to pay over \$1.60, we were, of course, interested in how long he could actually hold the weight. One final point was explained to him:

While we often feel that we are so tired that we cannot go on, this is not really true. One can rarely be so

tired as not to be able to continue for 30 seconds. Accordingly, I would like you to give me one-half minute's notice before you actually drop the weight.

Results

Table 4 gives the results for the nine Ss tested. All but one S in the wake state immediately exceeded hypnotic performance. This S held the weight for 6 min. 5 sec. in trance, a very remarkable performance, but in a subsequent wake state dropped the weight after only 3½ min. The exception demonstrates very clearly that it is necessary to ego-involve the person in the task and to convince him of his ability to do it. He reported that the seven minutes that had been given as an illustration of "average performance" had seemed so long, and his hand became so tired after three minutes that he felt convinced that he would be unable to come even close to the average, so therefore "why bother to try?" The next day the S was more carefully motivated and encouraged. He was then able to hold the weight for over 10 minutes.

Discussion

This experiment does not purport to prove that there is no increase in physical capacity in the trance state. Because of the motivating nature of the trance state, and the operational difficulty in obtaining equal motivational states, it becomes a technical impossibility to prove conclusively whether increased physical capacity is produced or not. The data, however, do show that the usually observed increase in performance of trance Ss may be accounted for by motivational difference.

From a theoretical viewpoint the reinterpretations to which this study had led seem most significant. As long as we believe physical capacity to be in fact increased by the simple expedient of the induced trance, it becomes necessary to look for the focus of the trance in something neurophysiological. If, on the other hand, we can understand the apparent increase in physical capacity observed during the trance state in terms of differences of motivation, we are then led to view hypnosis in psychological terms. It is clear that this study says nothing about why the trance tends to increase motivation nor does it even prove that this is so. It merely shows that adequate

motivation in the wake state leads to levels of performance equal or better than those found in the trance.

An objection that might be raised takes the form of the question as to what would happen if similar motivational techniques were used in the trance state to those in the wake state. But this question has little bearing on the essential point. If application of these techniques should produce a trance performance greater than the wake performance, it could be interpreted as the result of combined effects of ego-motivation and the postulated increased motivation associated with hypnosis. If, on the other hand, performance in trance were not greater it could be argued that the type of ego-motivation used is not germane to the trance state.

It may, finally, be argued that the *S* in the wake state is, in fact, still in hypnosis, since the same experimenter who induced hypnosis conducts the second phase. Perhaps *Ss* performed better in the wake state because of the demand characteristics of the experiment, i.e., my expectation that they should do so! It is not easy wholly to refute this argument. That all previous studies are open to the same criticism does not answer the question. The clinical observation that the *S* does not look, act, or feel in any way the same in the hypnotic part and the waking part appears much more relevant. Nevertheless, I hope sometime to repeat the study with the aid of another hypnotist who believes in "the power of hypnosis" and who, therefore, expects *S* to do better in hypnosis than in the wake state. If it were possible for me to enable *Ss* subsequently to exceed their hypnotic performance, it would go far toward removing this objection, of which I was aware during the collection of data. A different way to check the results would compare the performances of "fake" and "real" *Ss*, using the strategy developed in the section entitled Cues Implicit in an Experimental Design.

REAL VS. "FAKE" HYPNOTIC SUBJECTS

The "real-fake" technique, a method of enabling *Ss* to simulate hypnosis, was developed to demonstrate the effect of role-play artifact on trance behavior. Differences between the real and faking *S* that cannot be accounted for by the faking situation may be

viewed as characteristic of hypnosis. Some behavior shown by both groups may, of course, also be a true characteristic of hypnosis since the fact that someone is able to simulate a given type of behavior does not indicate that it is not genuine in the nonsimulating group. For example, that it is fairly easy to simulate the compulsive quality of the trance does not imply that this quality is not germane to hypnosis. However, assertions that volitional capacities can be transcended in hypnosis would seem to require showing that faking *Ss* cannot produce similar performances.

The present use of real and faking *Ss* in a blind design appears to offer several advantages. It permits a rigorous control, in terms of behavior, of inquiry procedures designed to elicit demand characteristics of experiments. In the faking situation, the variable assumed to be the cause of the behavior can be omitted. If such behavior still occurs, it can then be accounted for adequately by the implicit demands of the situation. In this respect the technique may have useful application in other areas of psychology. With respect to hypnosis itself, the technique permits a rigorous control of experiments that claim to demonstrate transcendence of volitional capacities. It also helps to eliminate many biases almost universally present in hypnosis research and throws into relief certain differences between the "real" and "faking" groups which can then be attributed to the hypnotic state. These differences may be highly germane to the essence of hypnosis and seem to have been obscured by the overwhelmingly impressive nature of the phenomenon.

The Faking Subject

The situation of the simulating *S* may be viewed as one where he can (a) contribute to research by pleasing one experimenter and (b) satisfy different other needs by fooling another experimenter. The possibility of simultaneously satisfying usually incompatible needs leads to a state of peculiarly high motivation. Thus, he is highly motivated to play the role of a hypnotized *S* and continues to do so even in the face of great discomfort, without, however, the alteration of consciousness and the subjective experience found in hypnosis.

It had been anticipated that there would be differences in the clinical appearance of the

two groups of Ss, both when going into trance and while in trance. Since the hand levitation technique (Wolberg, 1948) was used to induce trance, it was expected that the "faking" Ss would lift their hands in a manner clearly indicating active volition rather than permitting the hand to rise slowly and gradually in response to a subjective feeling of lightness. Similar differences were anticipated in respect to eye-closure. When the inability to bend the extended right arm was suggested, differences were expected in the S's attempts to do so. The S's responses to negative auditory hallucinations and positive visual hallucinations, analgesia and anesthesia, pseudo identities in the trance, and amnesia for the trance were all expected to be means of differentiating the two groups. The discovery that most of the simulating Ss did not differ behaviorally from those in deep trance came therefore with considerable surprise. True, it was possible to distinguish most of the "faking" group from the "real" group, but this discrimination was based on a clinical diagnosis with an inordinate amount of subjective uncertainty and about 20 % error. With prior knowledge that an S is simulating hypnosis it is very easy for the observer to discern cues that reveal the "faking" nature of the "trance" behavior. Without such prior knowledge, these cues are obscure and judgments can be made only with a great deal of subjective uncertainty.⁸ If it seems obvious to the experimenter that the S is faking, systematic error would again be introduced, and could affect the results. In such cases it seems desirable to discard the subject from the experiment, as was done in two instances in the present study.

Response to Pain in "Fake" and "Real" Subjects

Traditionally, hypnotists have viewed the induction of analgesia for highly painful stimuli as the best test for clearly distinguishing simulating Ss from the Ss in deep trance. The

author expected, therefore, to find that reactions to pain would distinguish between the "real" and "fake" groups.

The technique for producing pain for testing hypnotic analgesia which the author had been using routinely involves forcible flexion of the two terminal phalanges of the little finger and pressure of the experimenter's thumbnail on the dorsal aspect of the middle of the second phalanx of the S's little finger. These techniques failed to discriminate between the two groups. The "faking" group reported pain but indicated that they had had little or no difficulty in suppressing their reaction.

In order to provide a more painful, though still safe, stimulus a Harvard inductorium was employed. Standard Grass silver electrodes were applied to the sides of the index fingers and fixed with adhesive tape. It was found that the "faking" group flinched less and tolerated more pain than the "real" group. Actually, almost all of the Ss tolerated the maximal stimulation that could be applied by the inductorium.

Another procedure, not particularly painful in itself, was designed to maximize anxiety about pain. Negative auditory hallucinations were induced and at the same time an analgesia of one hand. A match was then lighted close to the S's ear and held under the third finger of this extended "pain insensitive" hand long enough to produce painful heat without causing a second degree burn. The experimenter tested the time interval on himself each time in order to determine what constituted a safe but painful stimulus. In no instance was an S ever subjected to a blistering burn. Again it was found that there were no noticeable differences between the simulating group and the deeply hypnotized Ss. However, postexperimental inquiry revealed that the simulating group experienced a high level of anxiety and some pain during the procedure, while those in trance felt little or no discomfort.

To summarize the results of these informal experiments with pain, no clear-cut behavioral differences between "real" and "fake" Ss have as yet been found, no matter what type of stimulus was employed. However, on postexperimental inquiries a very real difference was found in the S's attitude toward this part

⁸ The experimenter is currently attempting to study systematically the cues perceived as indicative of faking by an observer who knows that the S is simulating to see if consistent patterns appear. However, results thus far indicate that observers who had experience with observing faking Ss, knowing that they were simulating, did not thereby become more able to discriminate in the absence of such knowledge.

of the experiment and in his report of subjective pain.

Differences Between "Real" and "Fake" Subjects

The faking *S*'s condition is characterized by the presence of faking behavior and the absence of hypnosis. If we wish to regard true hypnotic behavior as the difference between what can be observed in faking and real *S*s, it is necessary to take into account the characteristics of the faking *S*s introduced by the act of simulating.

One of the observed differences is related to an attempt by the simulating group to avoid commitments whenever possible. In some aspects, the situation of the simulating group is analogous to that of the malingerer, although the motivation to mangle is different and usually stronger. The person who is simulating a disease attempts to give the appropriate behavioral and symptomological manifestations without subjectively experiencing the symptoms that would be present with the actual disease. The real hypnotic group is more akin to the hysteric who reacts to subjectively real symptoms without organic pathology. Rorschach studies on malingerers in the military service (Rosenberg & Feldberg, 1944; Ross, 1944) reveal a typical syndrome. The most significant aspect is a marked constriction of the record, with occasional bizarre responses, a feature that is understandable as an attempt to avoid giving oneself away in the face of an unknown situation while attempting to appear cooperative. Applied to the present situation, these features suggest a tendency for the simulating group to avoid responding whenever possible and when forced to respond in an unstructured situation to give the safest possible answer. Just this occurred. In the replication of the Ashley, Harper, and Runyon study, it was observed that the "faking" *S*s would tend to be more hesitant in giving a false life history after being told that they were rich or poor. This reluctance cannot be considered a particularly meaningful finding for the theory of hypnosis, since it should rather be attributed to the psychological situation analogous to malingering, and not to the presence or absence of hypnosis.

Differences observed which may be attrib-

uted to the presence or absence of trance were with respect to: (a) congruence, the appropriateness of a response to the apparent level of trance, and (b) trance logic, the apparently simultaneous perception and response to both hallucinations and reality without any apparent attempts to satisfy a need for logical consistency.

The "fake" *S*s tended to exhibit behavior that was incongruous with what clinically appeared to be the depth of trance. For example, an *S* who responded very slowly to hand levitation was given the suggestion that he could hear only my voice. This negative auditory hallucination was tested by rubbing the thumb against the first two fingers of the hand close to the *S*'s ear and asking him "do you hear anything but my voice?" A response congruent with the *S*'s behavior of slow hand levitation and in a medium trance would be either no response or a *slight* nod or shaking of the head. Frequently, a repetition of the question would be required to elicit even a nod of the head. An *immediate* and emphatic *verbal* denial of any auditory perception other than my voice was considered a typical incongruent response—incongruent both in respect to speed and modality of response.

The most reliable criterion, as it turned out, may be termed trance logic.⁹ It refers to the ability of the *S* to mix freely his perceptions derived from reality with those that stem from his imagination and are perceived as hallucinations. These perceptions are fused in a manner that ignores everyday logic. A good example of this process is the reaction of many *S*s in hypnosis who see a hallucination of a person sitting in a chair and describe it as "This is very peculiar, I can see Joe sitting in the chair and I can see the chair through him." This type of reaction, when made spontaneously, was absolutely diagnostic of the "real" *S*. Unfortunately, it did not occur spontaneously in every *S*; it was found, however, that "fake" *S*s *never* gave such a response.

In order to obtain the same type of data

⁹ This seems intimately related to primary process thinking or autistic thinking. To avoid the developmental, pathological, and theoretical implications of these concepts a new phrase seems preferable to describe the clinically observable phenomenon.

that had proved to be so characteristic of hypnosis, a special situation was devised. Throughout the experiments, one of the co-experimenters was in the room with the *S* and the author, taking notes. At the beginning of the experiment he sat to the side of the *S*, well within his visual field. Once the *S*'s eyes were closed, the co-experimenter rose very silently and walked behind the *S* out of the *S*'s visual field. Subsequently, the *S* was instructed to open his eyes while still in deep trance and a hallucination of the associate sitting in the chair was induced. This hallucination was facilitated by the fact that during the initial part of the experiment, just prior to the *S*'s closing his eyes, the associate was actually sitting in the chair. After the hallucination appeared to be fully accepted by the *S*, and he appeared to be responding to the hallucination as though it were actually the associate, he was instructed to turn around and look at the associate now behind him with the question, "Who is that behind you?" Almost invariably the "real" *S* would look at the associate, then quickly turn back to look at the empty chair, and back at the associate. After looking back and forth between the hallucinated and the "real" associate, the *S*s indicated verbally that they were perceiving two images of the same person.¹⁰ When asked about this, they tended to give bland responses such as "mirrors" or a "trick." Though they seemed somewhat confused, they did not become seriously disturbed.

The "fake" group, in most cases, either refused to see anyone behind them, or claimed that they could not recognize the person. Occasionally, they admitted recognizing the associate behind them and then claimed that the hallucination had vanished. (This reaction, while unusual among the real group, is occasionally found, especially if the *S* did not fully accept the initial hallucination.) Of approximately 30 "faking" *S*s, only two acted as if they saw two images of the same individual. The others, when asked during post-experimental inquiry about the reason for their response, gave a very significant answer. They stated that I had instructed them to

hallucinate, the associate in the chair and therefore, when I asked who the individual behind them was, they had to deny seeing him, or if they saw him, recognizing him. After all, there was only one such individual and I had already told them that they were supposed to see him sitting in the chair. This logical conclusion determined the response given by simulating *S*s; it did not occur to the overwhelming majority of the "real" *S*s who saw two images without any difficulty. This finding appears to represent a valid and significant difference. The "real" *S* responds to a subjectively real image of the associate in the chair. When asked about an objectively real image of the same individual, he is able to perceive this as well. He can respond to perceptions that are subjectively real and determined by the suggested environment, as well as to his actual perceptions of the real world, without attempting to satisfy a possible need to make them logically compatible. The absence of expression of a need for logical consistency seems, at this point, to be one of the major characteristics of hypnosis.

Because it is my belief that the "essence" of hypnosis will be found in the subjective experiences of the *S*, I have become increasingly interested in a series of techniques attempting to obtain data about the actual feelings and experiences of the *S*. In the future, I intend to develop inquiry procedures that will include "casual conversation" with another *S* who, in reality, is an investigator. It is hoped to elicit cooperation from the *S*'s friends. Casual preliminary attempts using such procedures indicate that material obtained in this way may be quite illuminating and not accessible to direct inquiry by the experimenter.

A POINT OF VIEW TOWARD HYPNOSIS

While much of the research described here appears to be explaining away the hypnotic phenomenon, the intention is rather to differentiate its valid and significant aspects from what might be termed artifact. One of the problems inherent in any study of hypnosis is that of its definition. There is high consensus of opinion about what constitutes hypnosis in terms of a variety of scales. However, the

¹⁰ This situation was originally discussed by Milton H. Erickson in a personal communication.

essential characteristics have remained obscure. A great many investigators have become impressed and fascinated by the apparent transcendence of normal physiological capacities in hypnosis. The present research program has made me increasingly skeptical of the experimental data that purport to support this view. However, clinical data obtained both by others and myself seem to show in a dramatic way that responses can be evoked in some *Ss* which they themselves could not perform voluntarily. Such phenomena seem to be limited to *Ss* who have a peculiar disposition in this direction. For example, authenticated cases of hypnotically induced blistering have been achieved only in individuals with previous dermatological histories.

One might hypothesize that the capacity to produce marked physiological alterations in hypnosis is confined to persons who have a readiness to somatize in the organ system being investigated, which will usually have been demonstrated by a history of similar pathology occurring spontaneously. Such findings do not preclude the possibility, of course, that a transcendence of normal volitional capacities in some areas may eventually be established in the laboratory as unequivocally due to hypnosis.

Aside from the controversial issue of such changes in physiological capacities, it appears that a universal effect of hypnosis on any *S* in deep trance can be delineated in terms of his subjective experience. Experience, after all, is not to be taken as an ephemeral or unimportant aspect of hypnosis but, rather, as extremely significant and, to the *S*, dramatic and striking.

Any *S* who has experienced deep trance will unhesitatingly describe this state as basically different from his normal one. He may be unable to explicate this difference, but he will invariably be quite definite and certain about its presence. Thus, one of the characteristics of hypnosis is that in deep trance the *S* experiences the state as discontinuous from his normal waking experience (though not always in the intermediate stages of trance). Hypnotic trance differs from pathological states, which may also be discontinuous, in that the *S* enters

and leaves the state in accordance with previously established "rules of the game."

Another aspect of this altered subjective state is one which the *S* describes as an inability to resist a cue given by the hypnotist. (Interestingly enough, if the *S* before entering the trance decides not to follow a specific suggestion he is able to resist it.)¹¹ The uniformity with which this compulsive quality is reported tends to make us accept it as a characteristic of hypnosis. However, it will not emerge as a difference in behavior between real and faking *Ss*.

Finally, an important attribute of hypnosis is a potentiality for the *S* to experience as subjectively real suggested alterations in his environment that do not conform with reality. In trance, the waking distinction between an imagined idea and what is perceived externally to the organism fades, and images may be perceived as originating from external reality. Thus, the waking individual, no matter how hard he tries to imagine that he saw someone sitting opposite to him, might at best be able to evoke some kind of imagery but would always be aware of the distinction between this and reality. The *S* in deep hypnosis may well be unaware of the distinction, though at some level he will always be able to discriminate.

In sum, the principal features of the hypnotic state are seen as changes in the subjective experience which are characterized by (a) discontinuity from normal waking experience, (b) a compulsion to follow the cues given by the hypnotist, (c) a potentiality for experiencing, as subjectively real, distortions of perception, memory, or feeling based on "suggestions" by the hypnotist rather than on objective reality, (d) the ability to tolerate logical inconsistencies that would be disturbing to the individual in the wake state.

SUMMARY

This paper has attempted to delineate some aspects of hypnotic phenomena which can be

¹¹ However, suggestions that are inconsistent with the basic "rules of the game" governing the implicit contract between hypnotist and *S*, as seen by the *S*, are, as a rule, not followed: e.g., antisocial and self-destructive acts, or any other suggestions running counter to basic ego needs or superego inhibitions.

rigorously tested and established. The hypothesis that the subject's "knowledge" regarding behavior in hypnosis influences his own hypnotic behavior was supported by an experiment. Students were exposed to a demonstration and lecture on hypnosis in which catalepsy of the dominant hand was mentioned as a common feature of trance behavior. Five out of nine volunteers exhibited this phenomenon under hypnosis. No students in a control group, who were given a similar lecture and demonstration but with no mention of catalepsy, showed the phenomenon.

An experiment performed by Ashley, Harper, and Runyon, which depends on hypnotic amnesia to explain the results, was repeated with the inclusion of a control group of subjects. These were subjects who simulated hypnosis but who were otherwise exposed to the same experimental situation as the hypnotic subjects. The behavior of the simulating group was indistinguishable from that of the "real" group, and both were indistinguishable from the results of the original study. Some doubt is thus cast on an explanation of the results in terms of hypnotic amnesia, and support is lent to the hypothesis that the demand characteristics of the experimental procedure may be a significant determinant of subject behavior.

In another experiment it was found that motivated subjects in the wake state held a weight at arm's length for a longer period of time than they did while in the hypnotic state. This result casts doubt on the notion that enhanced physical capacity is a primary characteristic of the trance state.

Differences between "real" and "fake" subjects were investigated. The major difference appears to be a tolerance by the "real" subject of logical inconsistencies.

It was concluded that in the absence of objective indices of hypnosis the existence of trance may be considered a clinical diagnosis. Until an invariant index of hypnosis can be established, such a diagnosis must be confirmed by the subject's report of alterations in his experience, since the real focus of hypnosis appears to lie in the subjective experience of trance.

REFERENCES

- ASHLEY, W. R., HARPER, R. S., & RUNYON, D. K. The perceived size of coins in normal and hypnotically induced economic states. *Amer. J. Psychol.*, 1951, **64**, 564-592.
- BINET, A., & FÉRE, C. *Animal magnetism*. New York: Appleton, 1888. (Trans.)
- BORING, E. J. *A history of experimental psychology*. (2nd ed.) New York & London: Appleton-Century-Crofts, 1950.
- BRUNER, J. S., & GOODMAN, C. C. Value and need as organizing factor in perception. *J. abnorm. soc. Psychol.*, 1947, **42**, 33-44.
- COUÉ, E. Self mastery through conscious autosuggestion. *N. Y. Amer. Library Serv.*, 1922, P. 83.
- DUMAURIER, G. *Trilby*. New York: Harper, 1895.
- DYNES, J. B. An experimental study in hypnotic anesthesia. *J. abnorm. soc. Psychol.*, 1932, **27**, 79-88.
- ESTABROOKS, G. E. *Hypnotism*. New York: Dutton, 1948.
- HULL, C. L. *Hypnosis and suggestibility*. New York & London: Appleton-Century, 1933.
- JENNESS, A. Hypnotism. In J. McV. Hunt (Ed.), *Personality and behavior disorders*. New York: Ronald, 1944, P. 470.
- LECRON, L. M., & BORDEAUX, J. *Hypnotism today*. New York: Grune & Stratton, 1947, P. 103.
- MCDUGALL, W. *Outline of abnormal psychology*. New York: Scribner's, 1926.
- MANN, T. *Mario and the magician*. Lowe-Porter (Trans.) New York: Knopf, 1931.
- MAYER, L. *Die Technik der Hypnose*. Munich: J. F. Lehmanns Verlag, 1951, Pp. 142-143.
- MOLL, A. *Hypnotism*. (5th rev. ed.) New York: Scribner's, 1904.
- NICHOLSON, N. C. Notes on muscular work during hypnosis. *Johns Hopkins Hosp. Bull.*, 1920, **31**, 89.
- ORNE, M. T. The mechanism of hypnotic age regression: An experimental study. *J. abnorm. soc. Psychol.*, 1951, **46**, 213-225.
- ORNE, M. T. Die Leistungsfähigkeit in Hypnose und im Wachzustand. *Psychol. Rdsch.*, 1954, **5**, 291-297.
- PATTIE, F. A., JR. The genuineness of hypnotically produced anesthesia of the skin. *Amer. J. Psychol.*, 1937, **49**, 435-443.
- PAVLOV, I. P. The identity of inhibition with sleep and hypnosis. *Sci. Monthly*, 1923, **17**, 603-608.
- ROSENBERG, S. G., & FELDBERG, T. M. Rorschach characteristics of a group of malingerers. *Rorsch. Res. Exch.*, 1944, **8**, 141-158.
- ROSS, Y. D. The use of the Rorschach method in the Canadian Army. *Rorsch. Res. Exch.*, 1944, **8**, 159-161.
- ROUSH, E. S. Strength and endurance in the waking and hypnotic state. *J. appl. Physiol.*, 1951, **3**, 404-410.
- SARBIN, T. R. Contributions to role taking theory: I. Hypnotic behavior. *Psychol. Rev.*, 1950, **57**, 255-270.
- SCHILDER, P. *The nature of hypnosis*. Gerda Corvin (Trans.) New York: International Universities Press, 1956.
- STOKVIS, B. *Hypnose in der Aertzlichen Praxis*. New York: S. Karger Basel, 1955.

- STRICKLER, C. B. A quantitative study of post-hypnotic amnesia. *J. abnorm. soc. Psychol.*, 1929, **24**, 108-119.
- WEITZENHOFFER, A. *Hypnotism*. New York: Wiley Press, 1953.
- WELLS, W. R. Experiments in "waking hypnosis" for instructional purposes. *J. abnorm. soc. Psychol.*, 1923, **18**, 389-404.
- WHITE, R. W. A preface to a theory of hypnotism. *J. abnorm. soc. Psychol.*, 1941, **36**, 477-506.
- WILLIAMS G. W. The effect of hypnosis on muscular fatigue. *J. abnorm. soc. Psychol.*, 1929, **24**, 318-329.
- WILLIAMS, G. W. A comparative study of voluntary and hypnotic catalepsy. *Amer. J. Psychol.*, 1930, **42**, 83-95.
- WOLBERG, L. R. *Medical hypnosis*. Vol. I. New York: Grune & Stratton, 1948.

Received January 22, 1958.

DRIVE STRENGTH AND SYMPTOM MANIFESTATIONS¹

J. R. WITTENBORN

Rutgers University

AND NATHAN S. KLINE

Rockland State Hospital

THIS report describes a study of the validity of a new procedure for inferring drive strength. The rationale for this procedure takes its origin in Hull's proposal that the reaction potential of a response may be regarded as a function of two intervening variables, drive strength and habit strength (Hull, 1952). Habit strength is the variable which accrues from learning, i.e., from the frequency with which a response has been reinforced, and reaction potential refers to the energetic aspects of the response, e.g., latency, duration, amplitude, etc. Hull proposed that if one could assume relatively small individual differences in the habit strength of the response, the individual differences in the reaction potential of a response could be due to individual differences in drive strength (Yamaguchi, 1951). If, for example, one were considering individual differences in the reaction potential of a response which is a highly practiced one for all members of the sample, one could infer that these individual differences would be primarily expressive of individual differences in drive strength.

In attempting to make practical use of this principle for inferring individual differences in drive strength, one must be prepared to assume that for the sample in question there are no important differences in the strengths of the habits involved. One must, moreover, be prepared to make distinctions in the quality of the responses in a way which represents differences in reaction potential (Hull, Felsing, Gladstone, & Yamaguchi, 1947).

The rating scales which we now seek to validate were constructed by the application of this principle. We were interested in inferring differences between individuals with respect to the strength of the drives involved in a given situation when one resists the re-

quests of one's superior, seeks the attention of a superior, or expresses pleasure. We assumed that the various acts which may be involved in these classes of behavior were sufficiently well learned that they entailed no major differences in habit strength and could, therefore, be primarily expressive of differences in drive strength. The next step required that we be able to distinguish between persons with respect to differences in the reaction potential of the various classes of responses under consideration. Since we wished to infer drive strength on the basis of behavior ratings, our use of the various approaches to differences in reaction potential was analogical and not directly specific (as is the custom in most laboratory studies). Nevertheless, the concept of reaction potential was interpreted in nine different ways:

1. *Threshold*, the intensity or the force of the stimulating situation which is required to elicit a response, is a familiar description for performance in both physiological and psychological laboratories.
2. *Latency* is the interval separating the probable eliciting stimulating situation and the appearance of the response in question.
3. The *amplitude* of a response usually refers to the force with which the response is expressed and directly implies energy manifestation per se. (It could also be construed to refer to the extensity of motion.)
4. The *speed* with which the response is executed is a quality which is commonly used in the psychological laboratory.
5. The *duration* of the response is a familiar quality for the description of a performance.
6. The *refractory* aspects of the behavior have to do with the probability of a given kind of performance occurring repeatedly as contrasted with the quality of behavior where the performance once elicited cannot be readily evoked again for some time. We surmised that the stronger the drive, the less likely a refractory phase.
7. *Stimulus generalization* is a quality of performance which repeatedly appears in the experimental literature of psychology and refers to the diversity of the situations which will elicit the performance in question. We expected that the stronger the drive for a given response, the greater the variety of stimuli which will elicit it.
8. A kindred concept is *response generalization*. This

¹ We gratefully acknowledge the assistance of John Blair of the Rockland Research Facility, and John Doebele, Ruth Raime, and Alfred Hall, who assisted in collecting and analyzing the data.

refers to the diversity of responses which may be elicited in a given stimulus situation. Our expectancy was that the stronger the drive for any one response, the greater the chance of similar responses.

9. The final quality has to do with *superfluity* of behavior, i.e., the diversity of responses which simultaneously occur in a given situation. In a given situation, some persons may express a diversity of response forms more or less simultaneously so that one would say that their behavior has a superfluous quality. In the same situation, another person might respond in one manner only.

It was possible to construct a rating scale for each of the nine interpretations of reaction potential. These nine rating scales were then applied to each of the three different classes of behavior under consideration: resisting superior, seeking the attention of superior, and expressing pleasure.²

In order to justify even a tentative interpretation that these rating scales, based on the various analogical interpretations of reaction potential, express individual differences in the strength of the drive involved in the behavior, it is necessary that the various rating scale interpretations for the reaction potential of a given class of behavior distinguish between patients in a similar way. The required pattern of intercorrelations was reported in an earlier publication (Wittenborn, 1957); in general, when resistance to the request of a superior was rated according to the nine different interpretations of reaction potential, the various rating scales tended to be consistently intercorrelated. The same pattern of consistent correlation was found for the scales applied to seeking the attention of a superior and to expressing pleasure.

Because of the rational considerations which determined the way in which these rating scales distinguish between individuals and because of the pattern of internal consistency among the rating scales, it was inferred that rating scales based on an analogical use of reaction potential *could* express individual differences in the strength of the drives contributing to the class of behavior under consideration. The present report is concerned with an effort to show that individual differences distinguishable on the basis of this kind of procedure *do* show a predictable rela-

tionship with certain independent indications of drive strength.

HYPOTHESES AND METHOD

It is commonly assumed that anxious patients have a low order of drive for the expression of pleasure. Depressed patients have a low order of drive for the expression of pleasure, and they are rarely motivated to initiate associations with others. It was further assumed that manic patients are characterized by drives to both resistive and attention-demanding behavior and that hysterics are likely to be particularly attention-demanding. Such implied behavioral differences among mental hospital patients can be summarized in quantitative form by the use of a set of symptom rating scales developed in the course of earlier investigations (Wittenborn, 1954).

The following hypotheses are stated for the purpose of examining the validity of the scales:

1. The various scales for inferring strength of drives for expressing pleasure would tend to be consistently related, but in an inverse manner, with the degree to which the patient's symptoms are characteristic of acute anxiety.

2. The various scales for inferring strength of drives for expressing pleasure would tend to be consistently related, but in an inverse manner, with the degree to which the patient's symptoms were depressive in nature.

3. The various scales for inferring strength of drives for seeking the attention of superiors would also be consistently related, but in an inverse manner, with the degree to which the patient's symptoms were depressive in nature.

4. The various scales for inferring strength of drives for seeking the attention of superiors would tend to be consistently related with the degree to which the patient manifested hysterical symptoms.

5. The various scales for inferring strength of drives for seeking the attention of superiors would also be consistently related with the degree to which the patient's symptoms were characteristic of the manic states.

6. The various rating scales for inferring the strength of the drives for expressing resistance to the requests of superiors would tend to be consistently related with the degree to which the patient's symptoms were characteristic of manic states.

For the purposes of the present study, the rating scales were applied to the behavior of 46 young patients at the Rockland State Hospital. These patients were carefully observed for a two-week period by student raters who were trained in the use of both the new scales

² Copies of the 36 rating scales are available upon request.

TABLE 1

CORRELATIONS* BETWEEN ACUTE ANXIETY CLUSTER SCORE AND REACTION POTENTIAL RATINGS

Ratings for Reaction Potential	Resistance to Requests	Expression of Pleasure	Seeking Attention
I Threshold	-.58	-.62	—
II Latency	—	.65	.65
III Amplitude	—	-.68	—
IV Rate	—	-.70	—
V Duration	—	—	—
VI Refractory	—	—	—
VII Stim. Gen.	-.79	-.68	—
VIII Response Gen.	-.55	-.79	—
IX Superfluity	-.52	-.78	—

* Significant at the .05 level.

Note.—These are tetrachoric r 's based on 46 young mental hospital patients.

TABLE 2

CORRELATIONS* BETWEEN DEPRESSED STATE CLUSTER SCORE AND REACTION POTENTIAL RATINGS

Ratings for Reaction Potential	Resistance to Requests	Expression of Pleasure	Seeking Attention
I Threshold	-.62	-.78	-.35
II Latency	.48	.82	.70
III Amplitude	-.32	-.62	-.59
IV Rate	—	-.75	—
V Duration	—	-.62	-.55
VI Refractory	—	.50	—
VII Stim. Gen.	—	-.75	-.57
VIII Response Gen.	-.35	-.82	-.48
IX Superfluity	—	-.80	-.48

* Significant at the .05 level.

Note.—These are tetrachoric r 's based on 46 young mental hospital patients.

developed for the purpose of inferring drive strength and in the use of the symptom scales. The ratings were made with unusual care, and the raters did not know the purpose of the ratings or the manner in which the data would be analyzed.³

RESULTS

As scrutiny of Tables 1, 2, 3, and 4 reveals, the data for our sample of 46 patients satisfied the requirements of the predictions. Our assumption that this new rating principle can yield distinctions which are expressive of differences in the strength of the patients' drives is not challenged by these data; as a matter of fact, the data suggest that such an approach permits us to infer differences in drive strength which may be plausibly related to other aspects of behavior expressive of differences in

³ The patients, adolescent and juvenile males, were observed for 6- to 8-hour periods each day.

TABLE 3

CORRELATIONS* BETWEEN MANIC STATE CLUSTER SCORE AND REACTION POTENTIAL RATINGS

Ratings for Reaction Potential	Resistance to Requests	Expression of Pleasure	Seeking Attention
I Threshold	.48	—	.42
II Latency	-.29	—	-.36
III Amplitude	.58	.28	.30
IV Rate	.54	—	.51
V Duration	—	—	.60
VI Refractory	-.68	—	—
VII Stim. Gen.	.52	—	—
VIII Response Gen.	.40	—	—
IX Superfluity	.34	—	—

* Significant at the .05 level.

Note.—These are tetrachoric r 's based on 46 young mental hospital patients.

TABLE 4

CORRELATIONS* BETWEEN CONVERSION HYSTERIA CLUSTER SCORE AND REACTION POTENTIAL RATINGS

Ratings for Reaction Potential	Resistance to Requests	Expression of Pleasure	Seeking Attention
I Threshold	.48	—	.62
II Latency	-.52	—	-.44
III Amplitude	.50	—	.46
IV Rate	.66	—	.61
V Duration	.44	—	.56
VI Refractory	-.70	—	-.58
VII Stim. Gen.	.52	—	.42
VIII Response Gen.	.46	—	.52
IX Superfluity	—	—	.52

* Significant at the .05 level.

Note.—These are tetrachoric r 's based on 46 young mental hospital patients.

drive. Certainly this seems to be a definite possibility for inferring some of the drive differences implicit in symptomatic behavior.

It will be noted that the tables do not include all the possible correlations. Tetrachoric correlation coefficients were determined for only those relationships which were found on the basis of preliminary chi square analysis to be significant at the .05 level. Since tetrachoric correlations were used, and since the sample is not large, no formal comparisons are made between the various correlation coefficients. It is not suggested that any particular significance be ascribed to any one of the correlation coefficients, and, according to our theory, we have no a priori reason for assuming that any one of our estimates of reaction potential is either a particularly valid or a particularly questionable basis for inferring drive strength. For the purposes of discussion, we shall ascribe some practical

significance where four or more of the estimates of drive strength are correlated with a symptom cluster.

Since the symptom cluster scores refer to familiar differences in patients and have some commonly accepted motivational implications, perhaps a discussion can best be organized about the symptom cluster scores. Table 1 shows, as predicted, that the most acutely anxious patients tend not to express pleasure. In addition to this predicted pattern of relationship, there is some indication that these patients are not impelled by strong drives to resist the requests of their superiors. In Table 2, we find that patients who are most depressed tend to be the least strongly motivated to express pleasure, and that these patients are also not strongly motivated to seek attention. In addition, there is some indication that these patients are not inclined to be resistive to their superiors. These trends are certainly plausible and are suggestive of the behavior patterns which are commonly observed to characterize the reactions of depressives. Thus, our new procedure reveals the similarity between depressed and acutely anxious patients: They are alike in the sense that drives toward the expression of pleasure are extremely weak and that drives to resist the requests of superiors tend also to be weak. Our procedure reveals a commonly noted difference between these two classes of patients; among the depressed patients, the drive to seek attention is distinctly weak, whereas among the anxious patients this drive seems neither characteristically weak nor strong. From such patterns of relationships it is apparent that there is a consistency between the strengths of the drives inferred from our procedure and the drive strengths implied by the patients' symptom manifestations.

In Table 3, we find as predicted that manic patients are characterized by drives to resist requests of superiors and by drives to seek the attention of superiors. Drives to express pleasure appear to be neither characteristically high nor characteristically low. It is interesting to observe in Table 4 that the drive pattern characteristic of manic patients is also characteristic of the patients who show the symptoms of conversion hysteria. The drives to seek attention are, of course, notoriously

characteristic of hysterical patients, and this fact served as the basis for the predicted relationship confirmed in Table 4. The drive to resist the requests of superiors was not predicted to be characteristic of the hysterical patients. Nevertheless, such motivation is implicit in the hysterics' use of physical symptoms as an excuse for not playing their proper roles and as a way of showing resistance to others, particularly to their superiors.

DISCUSSION

It should be noted that we use the concept of drive in terms of the energizing qualities of behavior in the tradition of the learning theorists. Drive is not used here in the psychoanalytic sense, and our procedure does not contribute to an analysis of intrapsychic conflicts. According to a tradition beginning with Woodworth and characterizing many experimental psychologists, drive refers to differences in the energizing aspects of behavior. This is not to deny the value of analyses of intrapsychic dynamics; it is suggested, however, that the drive strengths inferred by the use of the present procedure are the drive qualities which are expressible in overt behavior and not the strength of covert drives which remain in mutually suppressive dynamic balance without finding overt expression. In other words, if one can make a distinction between overt *potentiated* drive and covert dynamically *bound* drive, our exploration is an attempt to infer the strength of potentiated drive expressed in overt behavior. This approach is not considered to be simply a description of the energetics of the behavior alone, because a distinction is recognized between the energetics, i.e., the reaction potential, of behavior which is well learned and the energetics of behavior which is not well learned. We propose that differences in drive strength be ascribed to differences in the reaction potential of equally well learned responses only. If one can not assume that the behavior involves responses equally well learned among the members of the sample, one must assume that individual differences in various interpretations of reaction potential express differences in the degree to which the responses are learned (habit strength) as well as possible differences in drive strength.

The writers see certain advantages in an approach which scales behavior on the basis of distinguishable individual differences in reaction potential. First, the distinctions in behavior are not made on the basis of the degree to which the behavior approaches some culturally determined norm, medical ideal, or ethical value system. The distinctions can be made for such relevant practical behavior, but they are organized on the basis of the energetics, i.e., the reaction potential, of the responses. It is feared that studies based on behavioral distinctions which are generated by practical values may have a significance which is more pertinent to the form of our society than it is to behavior in general. This leads to the second possible advantage of such a system of distinctions. During the '20s and '30s, most studies in behavior were conceived from the standpoint of basic science, using the physical and biological sciences as models and borrowing many of their traditions of measurement. During the '40s, however, studies in behavior were strongly influenced by the current problems of society and its members, and many of the new developments in assessment or measurement expressed distinctions in terms of these immediate practical problems. Another shift in emphasis is now discernible. With the introduction of energizing and tranquilizing drugs as aids in the treatment of behavioral problems there appears to be a resurgence of interest in the biological approach to the study of behavior. Accordingly, the writers believed that many research interests could be served by the development of concepts and procedures in measurement which were relevant to socially significant behavior, such as symptom formation, but which provided distinctions between individuals by the use of principles which were not dependent on distinctions in our social values, e.g., by the use of distinctions in behavior based on an analogical employment of essentially physiological variables.

In any new approach, considerable experience and conceptual refinement is required

before we know either its limitations or how best to apply it. For some kinds of behavior the various indications of reaction potential may have a different expressive significance and may not, therefore, be similarly indicative of drive strength. Such possibilities must be explored before we can identify the situations in which the reaction potential principle may be properly applied to inferring strength of drive (Seward, 1955).

SUMMARY

This report indicates that a new principle for inferring the drive strength of human behavior has a practical relevance for external criteria. Estimates of the strength of drives among mental hospital patients for resisting the requests of superiors, seeking the attention of superiors, and expressing pleasure were found to bear an expected pattern of relationships with the symptom manifestations of these patients as gauged by the use of a standard set of psychiatric symptom rating scales. It is suggested that this principle for inferring drive strength may be usefully applied to researches concerned with psychopharmacological agents which have practical effects on behavior by mediating physiological changes.

REFERENCES

- HULL, C. L. *A behavior system*. New Haven: Yale Univer. Press, 1952.
- HULL, C. L., FELSINGER, J. M., GLADSTONE, A. I., & YAMAGUCHI, H. G. A proposed quantification of habit strength. *Psychol. Rev.*, 1947, **54**, 237-254.
- SEWARD, J. P. The constancy of the I-V: A critique of intervening variables. *Psychol. Rev.*, 1955, **62**, 155-168.
- WITTENBORN, J. R. *Psychiatric rating scales*. New York: Psychological Corp., 1954.
- WITTENBORN, J. R. Inferring the strength of drives. In M. R. Jones (Ed.), *The Nebraska symposium in motivation*. Vol. V. Lincoln: Univer. Nebraska Press, 1957. Pp. 191-259.
- YAMAGUCHI, H. G. Drive (D) as a function of hours of hunger (h). *J. exp. Psychol.*, 1951, **42**, 108-117.

Received July 10, 1957.

VERBAL RATE, CONTENT, AND ADJUSTMENT RATINGS IN EXPERIMENTALLY STRUCTURED INTERVIEWS¹

FREDERICK H. KANFER

Purdue University

A MAJOR obstacle in the examination of the psychotherapeutic process has been the problem of discovering those dimensions of analysis most relevant to behavioral changes as a function of therapeutic experience. Investigation of psychological and linguistic variables in interview behavior can broadly be divided into two clusters of approaches with some overlap in the methods usually employed. On the one hand, numerous studies have dealt with content analysis (Auld & Murray, 1955) or the categorization of information which the observer can extract from the subject's (S's) verbal behavior. The choice of categories varies widely and is mainly dictated by the purpose of the analysis and the theory from which it is developed. The schemas range from semantic categories which require little inference on the part of the observer to those based almost completely on inferences about S's emotional state or his dynamics. An example of the former is the use of the verb/adjective ratio (Balken & Masserman, 1940), whereas the latter is illustrated by Murray's (1938) need analysis or Bellak's (1954) interpretations of the TAT. The second group of approaches is represented by the study of structural speech characteristics. In this method, the dependent variable is a physical measure of speech behavior, such as rate or duration of utterances (Goldman-Eisler, 1954; Matarazzo, Saslow, & Matarazzo, 1956). The underlying theoretical basis rests on the argument that such differences in patterns of speech are related to other personality characteristics.

Those who argue for a structural approach to verbal behavior consider speech patterns to be stable personality characteristics and to be unaffected by mild changes in emotional attitudes connected with content (Chapple & Arensberg, 1940). Similarly, research on con-

tent variables has generally not attempted to deal with structural variables. It is quite reasonable to assume, however, that varied emotional states affect speech rate, for example, as well as content. Several divergent theoretical views have suggested the emotionality of the speaker as a major determinant of speech rate. Within Hull's framework (Benton, Hartman, & Sarason, 1955), anxiety is regarded as a drive state, and the level of verbal activity is related to manifest anxiety. On the basis of Skinner's approach, the properties attributed to anxiety also suggest changes in verbal output with increasing anxiety. In two earlier studies, Kanfer (1958a, 1958b) has shown that verbal rate increases as a function of shock-induced anxiety. In addition, clinical observers have repeatedly called attention to the patient's rate of speech as an index of his emotional state. In psychopathology, the direction of change in speech rate has been thought to depend on the patient's patterns of defenses. For example, blocking and increased verbal output may indicate a heightened emotional state, related to what is talked about.

The first step in integrating these approaches is to ascertain whether the content and structural dimensions show any correlation under varying emotional conditions. Although independent manipulation of these conditions is the eventual goal, a start may be made by identifying changes in content through controlling topics and by using the usual methods of content analysis to infer changes of the speaker's emotional state. Possible variations in structural variables may then be observed as a function of these changes.

The major purpose of the present study was to explore the relationship between verbal rate and rate variability and two content categories, topics and a content-inferred measure of anxiety. On the basis of the foregoing considerations, the following specific hypotheses were proposed:

Hypothesis 1. In a group of relatively homogeneous Ss, interindividual similarities due to cultural influences

¹ This study was supported in part by Research Grant M-963 from the National Institute of Mental Health of the National Institutes of Health, Public Health Service. The data were gathered while the author was on the faculty of Washington University, St. Louis.

may exist with respect to the extent to which talk about certain topics modifies emotional states. If the topics evoke similar emotional responses due to cultural attitudes, then differences in verbal rate would be expected between topics.

Hypothesis 2. In addition to the effects of various topics on all Ss, idiosyncratic experiences of Ss in relation to these topics should result in arousing different levels of anxiety among them. In an interview setting, the extent of anxiety aroused in individual Ss by various topics is difficult to evaluate directly. However, anxiety level may be assessed from the content of the protocol. On the basis of earlier findings (Kanfer, 1958a), it is hypothesized that there is a positive correlation between verbal rate and a content-inferred measure of anxiety. Secondly, intra-individual rate differences among topics are expected to show the highest rates on those topics on which the individual's anxiety-ratings are also highest.

Hypothesis 3. In clinical work it has been suggested that such behavioral variability as restlessness, uneven speech, etc., are associated with anxiety. The third hypothesis therefore states that greater variability in speech behavior is related to such indices of high anxiety.

METHOD

Subjects and apparatus. Ss were 26 students of both sexes from the author's class in abnormal psychology. All were nonveterans, 19-21 years of age, and single. Data from two Ss had to be discarded because the highly personal material revealed forced *E* to modify the experimental procedure in order to reassure the Ss and to safeguard them from further stress. Illegibility of tape-recordings further reduced *N* to 20.

Interviews were conducted in the author's office, which was familiar to the students. The apparatus consisted mainly of recording and timing devices. A microphone was in full view of *S*. A beep signal was fed directly into the recorder at 5-sec. intervals. A hidden knife-switch controlled the onset of the beep signal. A modified desk-clock was used which could be reset at the start of the interview and could be read with an error of less than three seconds. The clock faced *E* and was half hidden behind books. Precautions were taken to avoid interruptions during the interview.

Procedure. Ss were told that the experiment concerned a study on a personality test. All Ss were given Rorschach tests by graduate students and were then given appointments "to answer several more questions." When seated for the interview, they were given the following instructions:

This is a study in which we want to find out how well we can integrate our findings on a personality test and on a typical psychiatric interview of people who are relatively free of any mental illness. We will need as much information and as accurate an account as you can give us. Therefore, you can help us if you answer my questions by talking about everything which comes to mind when I raise a topic. Tell me about your attitudes and feelings and give whatever thought or example comes to mind. It is important that you omit nothing. You can talk freely, since this interview will remain confidential. I am going to record it since I cannot possibly write down all you say, but I will use a code number to identify the interview and no one will have access to any individual interview. After you have answered a few questions, I will discuss with you a little bit what the personality test indicated. Do you have any questions now?

Four warm-up questions about *S*'s school work and major interests were given in order to establish a pattern of continuous speech. With each succeeding question, *E* allowed longer time periods before asking another. When necessary, *E* explained that *S* had to continue talking about the topic and that *E* would not answer questions or ask any until each topic was exhausted. On the last question, three minutes were allowed. By that time, all Ss had learned that it was their responsibility to continue to talk, and all understood the instructions. After this warm-up period, the experimental session was begun.

The experimental interview consisted of five questions given to all Ss in the same order. After asking each question, *E* allowed three minutes of uninterrupted speech by *S*. The following questions were used:

1. Family Relationships (FR)
Tell me a little about how you are generally getting along with your parents.
2. Self-Confidence (SC)
Now I want you to tell me a little about yourself. Do you generally feel pretty sure of yourself in most situations?
3. Achievement (AC)
Have there ever been times when you wished you were able to do something which you couldn't do or weren't competent to do?
4. Sex (SX)
Do you generally feel that you are attractive to members of the opposite sex and that you get along well with them?
5. Emotional Maturity (EM)
Do you think that you are emotionally grown up, that you are emotionally mature and able to accept things?

At the end of each three-minute period or as soon thereafter as *S* paused, *E* ended the

question by using one of two prepared summary statements which rephrased the question and indicated a qualified affirmative or negative reply to *S*. For example, after the first question (FR), *E* said: "In general, then, you feel that you get along pretty well (or, sometimes do not get along too well) with your parents."

E gave several practice interviews in order to learn to minimize such cues as nods or changes in facial expression. Eight interviews were given for the purpose of testing the equipment and procedure and standardizing *E*'s behavior. None of the pilot interviews were used in the actual study.

Treatment of data. Duplicate typescripts of all interviews were prepared. All records were transcribed verbatim, including such speech errors as repetitions, false starts, etc. Only vague sounds such as "eh" and "ah" were omitted. All punctuation was omitted. One copy was prepared for measuring speech rate and variability by marking the typescript at 5-sec. and 30-sec. intervals in accordance with the taped beep signal. Partial words, repetitions, and speech errors were counted as separate words. In view of the great variability of the 5-sec. units, all measures were obtained from 30-sec. samples.

For the evaluation of anxiety from content, a series of adjustment scales was constructed. *E* and two clinicians devised a separate five-point scale for each of the five topics.² The clinicians, who later rated all interviews, had no knowledge of the purpose of the study. Specifically, they were unaware that "structural measures" for the speech samples were available. The rating scales were anchored by supplying descriptive statements, specific to each topic, for each of the scale values. The raters agreed to deal with the interview content on a manifest level, i.e., their ratings were based only on such inferences which could be made directly from the data level and could be supported by specific portions of *S*'s content. Since most theoretical views hold that anxiety relates to adjustment and that good adjustment refers to a state of low anxiety which does not interfere with a person's effectiveness,

a general view of the content was taken by instructing the judges to rate adequacy of adjustment. The concept of adjustment was essentially defined with respect to *S*'s social effectiveness. Thus, such manifestations as anxiety or defensive operations aimed at anxiety-reductions (including symptoms), lack of realistic, flexible, or tolerant self-appraisal, and unrecognized conflicting behavior and attitudes or failure to attempt their resolution were considered indicative of poor adjustment. Ratings were assigned according to the frequency and intensity of such indices in the interview material. The rating scale was based on absolute criteria, so that no specific distribution of scores was expected for the population. After jointly rating two interviews from the pilot study, each judge independently rated the protocols. To reduce contamination between topics, each topic was rated for all *S*s in the order of topics, yielding one score for each *S* on each topic and a total adjustment score for each *S*, obtained by summing the five topic scores.

Two raters were used to obtain a measure of rater reliability. The rho of .75 between the two judges for the summed adjustment scores of each *S* indicated a reasonable reliability in judges' ratings of content. Rank-order correlations for individual topics ranged from .71 to .77. These coefficients indicate that judges' agreement did not vary greatly from topic to topic. Scores ranged along the entire scale. The mean adjustment ratings for topics varied from 2.6 to 2.8, indicating that no rater bias existed for any one topic. For statistical treatment, the ratings of the senior rater were used.³

RESULTS

Verbal rate. *S*s showed considerable variability and long pauses during the last-minute interval. In order to eliminate differences due to fatigue or lack of responses for a given topic, only the first four half-minute intervals were used in the analysis of verbal rate. Figure 1 presents the mean rates on all topics in half-minute periods. The analysis of variance summary in Table 1 indicates that the data do not confirm the hypothesis concerning the extent

² The writer wishes to acknowledge the contributions of G. J. Williams and R. A. Craddock, who served as raters.

³ The writer is indebted to R. C. Bilger for his valuable suggestions concerning the statistical analysis of the data.

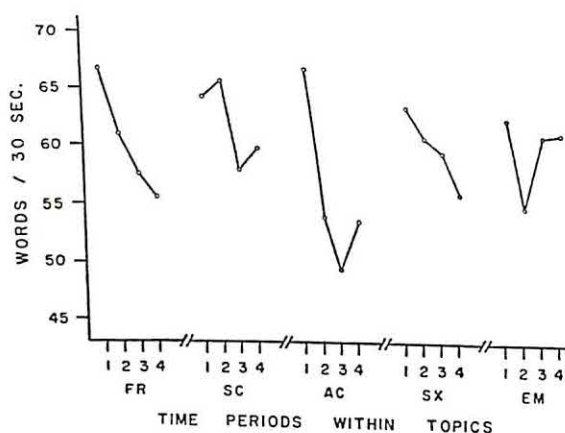


FIG. 1. VERBAL RATE WITHIN TOPICS IN 30-SEC. INTERVALS

of communality among *Ss* with respect to topics. Even among the relatively homogeneous group used in this study, no group differences in speech rate were observed on the five topics. However, if the triple interaction term is considered an estimate of error variance, then differences in verbal rate due to *Ss*' differential responsiveness to the various topics can be tested by $T \times S/P \times T \times S$. As shown in Table 1, the resultant *F* is significant at $p < .01$. Therefore, some of the rate variance can be attributed to the different effects which the topics had on individual *Ss*. Over periods, speech rate decreased significantly, Fig. 1 showing the decrements by topic. As each new topic was begun, *Ss* increased in their rate of talking. The *F* ratio for the $P \times T$ interaction shows no basis for inferring that the pattern of rate change within topics differed significantly among topics.

Adjustment ratings and verbal rate. Although the analysis of verbal rate indicated no differences between topics when groups were compared, it was hypothesized that the speech rate of individual *Ss* is related to a measure of their anxiety level inferred from their replies to each topic. The following correlation coefficients between verbal rate and adjustment ratings were obtained: On FR, $r = +.467$; on SC, $r = -.099$; on AC, $r = -.070$; on SX, $r = +.376$, and on EM, $r = -.146$. High ratings represented poor adjustment. The *r*'s were significant at $p < .025$ and $p < .05$ on topics FR and SX, respectively, by one-tailed test. On these topics, high rate of speech was significantly related to poor adjustment. As a

TABLE 1
ANALYSIS OF VARIANCE OF VERBAL RATE DURING 30-SEC. PERIODS

Source	df	MS	F
Subjects (<i>S</i>)	19	2452.87	
Topics (<i>T</i>)	4	389.40	<1
$T \times S$	76	420.39	1.75**
Periods	3	1284.15	4.01*
$P \times S$	57	320.25	
$P \times T$	12	239.41	<1
$P \times T \times S$	228	239.78	

* $p < .025$

** $p < .01$.

further test of the hypothesis, a sign test was run on the differences in rate between each *S*'s topics rated best and poorest in adjustment. Of the 20 *Ss*, 15 showed a higher verbal rate on topics rated poorest in adjustment, yielding a p of .042. Thus, *Ss* seem to talk significantly faster on those topics in which content suggests poorer adjustment than on those on which they are judged to be well adjusted.

Ss were then divided into two groups. All *Ss* whose total adjustment ratings fell above the median constituted the "good adjustment" group. Those below the median constituted the "poor adjustment" group. An analysis of variance of the verbal rate data failed to show differences between the two groups. Furthermore, topics also yielded a nonsignificant *F* ratio. Thus, no additional information was gained by this analysis.

Variability of verbal rate and adjustment ratings. A variability measure was obtained by subtracting the lowest rate in each two-minute sample from *S*'s highest rate on the same topic. A rank-order correlation for variability scores, summed over all topics, and total adjustment ratings yielded a rho of .021. Thus, no relationship was found between general adjustment and variability of speech rate. However, correlations between variability and adjustment ratings on each topic yielded the following *r* values: FR, .396; SC, .389; AC, .263; SX, .207; and EM, .162. The values for FR and SC are expected to occur by chance less than once in 10 times. Thus, poor adjustment tended to relate to high variability on the initial two topics. As the interview proceeded, variability showed a decreasing relationship to adjustment.

A coefficient of concordance was run to test the consistency of variability from topic to topic. The resultant chi square of 33.92 is significant at the .02 level. Therefore, variability rankings within the population of Ss used here show consistency from topic to topic. The results suggest, then, that the extent of variability of rate may be a fairly consistent individual characteristic.

DISCUSSION

The results show several general trends concerning verbal rate changes. First, rate declined considerably after S's initial response to a topic. However, introduction of a new topic restored verbal rate approximately to the level originally found on the preceding topic. This consistent recovery suggests that no general adaptation occurred during the interview. If speech rate is related to anxiety, then the implications of these results suggest a rearousal of anxiety with each new topic over the interview.

Reflecting on the first hypothesis, one finds that the topics show no similar patterns of effect on rate for all Ss. Thus, the assumption of a culturally determined, universal anxiety-arousing effect of the various topics is not borne out. When the adjustment ratings are used as criteria of anxiety and are derived from each S's response to specific topics, the observation of significant rate increases from "best" to "poorest" topics within Ss essentially confirms the second hypothesis. This finding is further supported by the significant $T \times S$ interaction, indicating that Ss' verbal rates varied with topics, but that the topics differed in the emotional impact which they had on different Ss.

With respect to the direction of change, the data show a facilitating effect of "anxiety" when this construct is related to adjustment ratings. This effect is in the same direction as observed earlier in a very different verbal situation (Kanfer, 1958a). It should be noted that S's verbal productivity in both situations was relatively free of demands for high quality. The facilitating effect on quantity of verbal output does not preclude concurrent changes in quality (e.g., appropriateness with respect to a problem) which the present study did not attempt to measure. For example, increased

redundancy and other content changes may occur. If they were found, they would suggest that emotional states may facilitate those aspects of verbal behavior which demand a lesser degree of organization while inhibiting more complex speech processes. Nevertheless, other factors may also contribute to variations in speech rate. The relatively low correlations obtained here preclude any use of the relationships for diagnostic purposes until a greater portion of the variance can be accounted for.

It is interesting to note that the highest correlations between rate and adjustment ratings were obtained on topics FR and SX. In comparison with the other topics, FR and SX differ in that they deal with areas in which verbal stereotypes or socially desirable "answers" are not readily available. In addition these topics required S's evaluation of his own behavior in specifically defined interpersonal situations, whereas the others did not. Finally, FR and SX deal with attitudes in areas which are considered crucial by most personality theories. Similar results were obtained by Moore, Soderberg, and Powell (1952), who found that stuttering increased during the topic "parents." If stuttering is viewed as a manifestation of anxiety (Wischnor, 1950), then Moore's results and the present ones would suggest that anxiety is more easily aroused in the parental and the sexual areas, and adjustment differences become sufficiently dominant to affect such behaviors as verbal rate and stuttering.

Observation based on the content analysis revealed that most Ss seemed to run out of relevant material on the remaining three topics and restructured the questions by relating specific incidents or elaborating some earlier statements. On FR and SX, however, Ss continued their self-evaluation with respect to the topic and no such restructuring took place. Their tendency to stay with the topic may be reflected in the continued decline of rate during the last period.

In a study which derives an hypothesis similar to the present one from Hull's theory and uses Taylor MAS scores as indices of anxiety, Benton, Hartman, and Sarason (1955) suggested that their failure to obtain a faster rate in their high anxious group was due to Ss' slowing down as their output increased. This

suggestion is verified by the present findings. The present results also point to Benton's use of a general criterion of anxiety as a possible source of failure to obtain rate differences between groups.

The data failed to establish a general relationship between speech variability and adjustment ratings. The higher correlations on the first two topics and their progressive decline suggest that poorly adjusted Ss are more variable in speech at first. Such adjustment-related differences did not persist over the interview.

The present study presented several problems in measuring rate. Verbal rate can be measured in any desired interval. However, the interval determines the specific definition of rate. A long interval masks such information as distribution and duration of pauses, "bursts" of verbal output, length of utterances, etc. On the other hand, a very short interval yields an extremely unreliable measure of rate which is strongly influenced by language structure. Although all these factors have been found to yield valuable information (Goldman-Eisler, 1954; Kelly & Steer, 1949), they cannot be investigated concurrently. The present measure was adopted after a preliminary study indicated that the 30-sec. interval yields relatively good stability for a measure of verbal units per time sample. The present rate measure can reveal changes in verbal output, but it does not reveal whether speed of utterance of a unit or pauses between units or other factors determine this change.

The data reveal the high sensitivity of the rate measure for individual response patterns in the experimental population. Thus, they confirm earlier research by Goldman-Eisler (1952) and others and support the contention that structural analysis of speech may be relevant to the investigation of personality. However, they also suggest the interrelationship of these physical measures with content-derived measures. The present study utilized only a very restricted range of adjustment ratings of a homogeneous population. The relationship between verbal rate and additional dimensions of adjustment needs to be explored further in populations representing a wider range of behavior patterns.

The adjustment measures admittedly represent an imperfect criterion variable for S's

"emotional state." They essentially reflect a clinical judgment criterion. Complete independence of the rate and adjustment measures cannot be proved despite attempts to avoid contamination of the adjustment ratings. Use of an external criterion (e.g., membership in a hospitalized or outpatient group or membership in a group whose emotional state is temporarily modified by experimental manipulations) remains a much more desirable step and is planned in further research.

SUMMARY

Twenty Ss from a selected, relatively homogeneous college population were given an experimental interview which they considered related to an earlier personality test. The interview was structured to allow collection of two-minute samples of uninterrupted speech, in reply to five questions relating to Family relationships (FR), Self-confidence (SC), Achievement (AC), Sexual attitudes (SX), and Emotional maturity (EM). Verbal output was tape recorded. Verbal rate was measured for 30-sec. intervals. Two clinicians rated the content of S's reply to each question separately on a five-point scale devised for this study. Rate variability was measured by subtracting each S's lowest rate from his highest rate on the same topic. There was no evidence that mean verbal rate differed among topics. However, speech rate decreased significantly within each topic. Furthermore, topics showed differential effects on the rate of Ss, as indicated by a significant topic \times subject interaction. Significant correlations between high rate and poor adjustment ratings were obtained for topics FR and SX. In addition, it was found that Ss talked significantly faster on those topics on which their adjustment was rated poorest when compared with topics on which they were rated as well adjusted. High variability of verbal rate showed a moderate correlation, with poor adjustment on the initial two topics.

The results were related to a general hypothesis of the facilitating effect of anxiety on verbal rate, and the present findings were considered as suggestive of relationships between verbal rate, topics, and those aspects of S's emotional state (anxiety) which overlap with a clinical evaluation of his adjustment derived from his verbalization.

REFERENCES

- AULD, F., JR., & MURRAY, E. J. Content-analysis studies of psychotherapy. *Psychol. Bull.*, 1955, **52**, 377-395.
- BALKEN, EVA R., & MASSERMAN, J. H. The language of phantasy: III. The language of the phantasies of patients with conversion hysteria, anxiety state and obsessive-compulsive neurosis. *J. Psychol.*, 1940, **10**, 75-86.
- BELLAK, L. *The TAT and CAT in clinical use*. New York: Grune & Stratton, 1954.
- BENTON, A. L., HARTMAN, C. H., & SARASON, I. G. Some relations between speech behavior and anxiety level. *J. abnorm. soc. Psychol.*, 1955, **51**, 295-297.
- CHAPPLE, E. D., & ARENSBERG, C. M. Measuring human relations: An introduction to the study of the interaction of individuals. *Genet. Psychol. Monogr.*, 1940, **22**, 3-147.
- GOLDMAN-EISLER, FRIEDA. Individual differences between interviewers and their effect on interviewees' conversational behavior. *J. Ment. Sci.*, 1952, **98**, 660-671.
- GOLDMAN-EISLER, FRIEDA. On the variability of the speed of talking and its relation to the length of utterances in conversations. *Brit. J. Psychol.*, 1954, **45**, 94-107.
- KANFER, F. H. Effect of a warning signal preceding a noxious stimulus on verbal rate and heart rate. *J. exp. Psychol.*, 1958a, **55**, 73-80.
- KANFER, F. H. Supplementary Report: Stability of a verbal rate change in experimental anxiety. *J. exp. Psychol.*, 1958b, **56**, 182.
- KELLY, J. C., & STEER, M. D. Revised concept of rate. *J. Speech Dis.*, 1949, **14**, 222-226.
- MATARAZZO, J. D., SASLOW, G., & MATARAZZO, RUTH. The interaction chronograph as an instrument for objective measurement of interaction patterns during interviews. *J. Psychol.*, 1956, **41**, 347-367.
- MOORE, W. E., SODERBERG, S., & POWELL, DONNA. Relations of stuttering in spontaneous speech to speech content and verbal output. *J. Speech Dis.*, 1952, **17**, 371-376.
- MURRAY, H. A. *Explorations in personality*. New York: Oxford, 1938.
- TAYLOR, JANET A. Drive theory and manifest anxiety. *Psychol. Bull.*, 1956, **53**, 303-320.
- WISCHNER, G. J. Stuttering behavior and learning: A preliminary theoretical formulation. *J. Speech Dis.*, 1950, **15**, 324-335.

Received December 18, 1957.

HOSTILITY IN RORSCHACH CONTENT AND OVERT AGGRESSIVE BEHAVIOR¹

ALAN P. TOWBIN²

West Haven Veterans Administration Hospital and Yale University

IN VIEW of the evidence (Calden & Cohen, 1953; Gibby, 1951; Kimble, 1945; Klatskin, 1952; Lord, 1950; Sanders & Cleveland, 1953; Singer, Meltzoff, & Goldman, 1952) that Rorschach responses vary with situational factors, it seems reasonable to regard the test responses as a function of the testee's perception of the test situation, rather than as a function of his perception of the test stimuli alone. The general form of the relation between test responses and perceived situation is a matter of practical as well as theoretical interest, since it may be necessary to take these situational effects into account in interpreting the test data. Several writers (Joel, 1949; Miller, 1953; Sarason, 1950; Schachtel, 1945) have suggested that the test protocol may be regarded as a sample of behavior in an interpersonal situation, that is, behavior in a situation characterized by the mutual awareness and influence of the two participants. This conceptualization suggests the hypothesis that test responses are influenced by the perceived situation in the same way that the perceived situation influences behavior in other interpersonal settings. Furthermore, it suggests that given comparable sampling conditions, Rorschach responses and interpersonal behavior are positively correlated.

In the experiment reported here, the category of overt behavior studied was *aggression*, overt hostility directed at some object other than the person of the aggressor. It is a commonplace observation, as well as a theoretically important one (Dollard, 1947, pp. 269-281), that aggression tends to be inhibited as a

function of the perceived capacity of the object of aggression to influence the condition of the aggressor. This capacity of the object is here designated *power*. *Status*, an individual's position in an institutional hierarchy, is closely though imperfectly correlated with his actual power over other members of the institution. Based on these considerations, the following predictions were made.

1. A positive correlation obtains between Rorschach test measures and independent behavioral measures of aggression.

2. The tendency of testees to produce aggressive Rorschach responses is inversely related to the perceived power of the examiner.

3. The tendency of testees to produce aggressive Rorschach responses is inversely related to the perceived status of the examiner.

METHOD

Design

The criterion measure of overt aggression was a history of such activity over a four-month observation period. S's assignment to an assaultive (A) or a non-assaultive (NA) group was based on the likelihood of his manifesting aggression. Ss in these two groups (A and NA) of schizophrenic patients were assigned at random to one of five conditions of Rorschach examination. Four conditions were procedures for influencing S's perception of E's power and status. The fifth condition followed routine hospital testing procedure and served as a control for the effects of the four experimental conditions. Rorschach records were scored for several categories of hostile content, only one of which, aggressive content (H₁), is relevant to the hypotheses tested here.

Subjects

Ninety-six male patients, diagnosed schizophrenic reaction (Committee on Nomenclature and Statistics of the American Psychiatric Association, 1952), were selected from the population of a large mental hospital. All were of at least average intelligence, under 50 years of age, and with one exception, had no diagnosed organic brain syndrome. The NA group was matched to the A group with respect to age, race, years of school attendance, diagnosis, interval since last Rorschach examination, and years of illness.³

¹ Based on a dissertation submitted in partial fulfillment of the requirements for the degree of doctor of philosophy in Yale University. The writer is indebted to and here thanks Seymour Sarason, Irving Janis, and Harold Kelley for their guidance in the execution of this study.

² This research was carried out while the writer was at F. D. Roosevelt Memorial V.A. Hospital, Montrose, N. Y. The writer wishes to acknowledge the support and the help of the psychiatric and psychological staff and psychology trainees who participated in this study.

³ A detailed description of the selection, matching, and experimental procedures and of the scoring systems used may be found elsewhere (Towbin, 1955).

The A group ($N = 48$) was composed of patients who, over a four-month period, had two or more assaultive episodes recorded in the daily reports of the ward nurse. No NA patient appeared on the daily report during this period for either assaultive or self-destructive acts. To qualify as an assault, the behavior had to be of such severity that another person was struck or hospital property damaged, or it was reported that forcible restraint of the angry patient was thought necessary to prevent such outcomes. Thus, the group assignment of S defined his status with respect to a tendency to aggression in his overt behavior.

Procedure

Immediately before testing, E and a confederate (C) carried on a brief discussion in S 's presence, the purpose of which was to influence S 's perception of E 's power and status. From the patient's point of view, the most powerful person in the hospital environment is probably the ward chief, a psychiatrist or psychologist who daily makes administrative and therapeutic decisions on matters of importance to patients. In the High Power condition, C was the patient's ward chief. In his talk with E , C indicated that his decisions about patients were influenced by E 's test reports. The patient was thereby given reason to believe that what he said during testing session could influence decisions affecting his interests. In the Low Power condition, C s were trainees in the Clinical Psychology Service of the hospital and were strangers to the patients. The Low Power interaction between E and C in the presence of S contained reference neither to the uses of test reports nor to E 's influence.

Higher status is presumably associated with greater power over those in lower statuses in the same section of the hierarchy. High rank in one section of the hierarchy does not necessarily carry with it power over those in lower ranks in other sections. Since the symbols of high status are often similar from one section to another, e.g., terms of deference, a high status person may elicit inhibition of aggression in an individual in the absence of indications that he has real power over the individual. In the interactions between E and C , many cues to their status relation are conveyed in the High and Low Status conditions.

The interactions between E and C were as follows:

1. *High Power, High Status:* The power figure (C) conveys to S that E is a colleague and that E 's opinions, based on test results, contribute to his (C 's) decisions.

2. *High Power, Low Status:* The power figure (C) presents E to S as his assistant; it is implicit in the interaction between C and E that E is doing testing at C 's direction and reporting to him.

3. *High Status, Low Power:* C presents himself as E 's assistant; the interaction between E and C follows the course of the second condition with the statuses of E and C reversed and with neither having power in relation to the patient.

4. *Low Status, Low Power:* C interacts with E to convey to S that E and C are equals, neither of whom has power in relation to S .

5. *Control:* No confederate was used and the procedure usually employed at the hospital in intro-

ducing patients to the examination situation was followed.

The scoring system of Elizur (1949) and the results of other investigators (Gorlow, Zimet, & Fine, 1952; Sanders & Cleveland, 1953; Walker, 1951) using that scoring system were helpful in the development of the scoring procedure used here. An *aggressive Rorschach response* (H_1) was one in which hostility was directed from the testee to the card, whether the explicit object of the attack was a percept, the blot, or the test as a whole. H_1 was scored for attacks on the test ("This is crazier than I am!"), derogations of the blot (weird, messy), attribution of derogatory qualities to percepts (ugly, servile), blaming the card for S 's failure to respond, and pointing out faults or inadequacies in the blot or percept.

A second category called H_2 involved no activity. It was simply scored for objects with a customary aggressive function, e.g., gun, teeth, etc. Finally, hostility from card to testee (aggressive figures like witches or giants), from an unspecified source to the blot (percepts described as the objects of attack, as injured, etc.), and hostility within the card (two figures duelling), were all scored H_3 . All responses received equal weight. S 's H_1 score was simply the number of Rorschach responses categorized H_1 .

Twenty Rorschach records were drawn at random from the 96 cases, and one copy of each was rescored independently by two clinical psychologists in addition to E .⁴ Since H_2 scores occurred relatively infrequently, they were combined with H_3 scores in computing the correlations between the three sets of scorings. The correlation coefficients for H_1 were .82, .95, and .72; the mean coefficient was .86. The coefficients for the combined H_2 and H_3 scores were .75, .89, and .81; the mean coefficient was .83. Thus, the scoring procedure showed an acceptable degree of reliability.

After the examination, S was asked his idea of the reason for the examination. S was considered to regard the test as consequential if his response mentioned an idea that the testing had to do with treatment, with evaluation of positive or negative personal qualities, with a decision affecting S , or if he said he thought someone had directed E to test him.

The interview records of two cases from each of the ten cells of the design were drawn at random, shuffled, and combined. These records were used to devise a scoring procedure likely to encompass the range of responses given by S s, but were not used for any item selection. These 20 cases were then returned to the remaining pool of protocols, and all interview records were scored in the same way. Twenty interview records drawn at random from the 96 were turned over to two clinical psychologists for rescoring.⁵ Pairing the categorizations made by the investigator and the independent scorers gave 17 out of 20 possible agreements. Thus, the reliability of the categorization seemed satisfactory.

Only the first three responses to each Rorschach card were scored in order to control the variance in R .

⁴ The writer wishes to thank Harold Wilensky and Jerome Singer for participating in this part of the study.

⁵ Pearl Greenberg and Renata Calabresi carried out this rescoring, for which the writer here thanks them.

TABLE 1
SUMMARY OF ANALYSIS OF COVARIANCE:
 $X = R, Y = \log (H_1 + 1)$

Source of Variance	X			Sum xy	Y			Corrected Y		
	df	ms	F		df	ms	F	df	ms	F
Treatments	4	10.7	—	1.27	4	.055	—	3	.046	—
Criterion	1	86.3	2.91	13.04	1	2.18	27*	1	1.90	26*
Interaction	4	7.8	—	1.55	4	.023	—	3	.022	—
Error	86	29.7	—	36.31	86	.080	—	85	.073	—
Total	95			50.72	95			92		

* Significance of .01 or above.

To estimate the degree of association between R, and the various measures of hostility (the criterion, H₁ scores, and H₂-plus-H₃ scores), the relevant distributions of scores were tabulated in two-by-two tables and chi square calculated in each instance. R was uncorrelated with the criterion of membership in the A or NA groups but correlated with H₁ and H₂-plus-H₃ scores to a degree significant beyond the .01 level. Consequently, in the tests of the experimental effects, the variance attributable to R was partialed out by means of the covariance technique.

The predictions stated earlier may now be put in more specific form:

- 1. The production of H₁ responses is positively correlated with membership in the A group.
- 2. Ss examined under the High Power condition produce fewer H₁ responses than testees examined under other conditions.
- 3. Ss examined under the High Status condition produce fewer H₁ responses than testees examined under other conditions.

RESULTS

Category H₁ includes both perceptions reported by the testee and his remarks and comments about the test, the card, etc. In order to test Prediction I with respect to both reported perceptions and remarks, separate analyses were carried out for each kind of behavior scored H₁ and for the two combined. The distributions of these scores were dichotomized and entered in a fourfold frequency distribution against group membership. The significance of the association was estimated by means of the chi square technique and phi coefficients were calculated as an estimate of the degree of relation. For H₁ over all, chi square was 26.33; phi was .52. For *percepts* scored H₁, a chi square of 24.14 and a phi of .50 were obtained. For *remarks*, chi square was 6.74; phi was .26. The relationships were all significant at the .01 level or beyond. The relation between the two subcategories of H₁ was sig-

nificant beyond the .01 level; chi square was 9.09 and phi .31.

These results indicate that it is possible to identify behavior in the Rorschach protocol and in an extra-test sample which can be ordered to the same class of events. The separate analyses indicate that the relation between H₁ and the criterion is not attributable solely to the testee's admittedly social behavior, i.e., remarks addressed to E: the customarily scored "percept" is also positively correlated with the criterion. The significant correlation between the two subcategories supports the procedure of combining them in H₁.

In the test of the effect of experimental treatments on the H₁ response (Predictions II and III), the logarithmic transformation was applied to the H₁ scores. Bartlett's test of homogeneity of variance yielded a chi square of 14.224, with a confidence level of .11. Hence, analysis of covariance was applied to these data. The results of the over-all analysis of covariance are presented in Table 1.

None of the results of the over-all analysis of covariance were significant except the criterion group difference, which was significant beyond the .01 level. This is no more than another description of the high correlation between H₁ and the criterion.

Only the difference between the High Power and the other treatments is pertinent to Prediction II. The adjusted mean (Cochrane & Cox, 1950, p. 75) of the High Power group was .417; that of the Low Power-plus-Control groups was .503. The difference failed to reach an acceptable level of statistical significance ($t = 1.52, p = .064$), and the null hypothesis cannot be rejected.

The difference between the adjusted means of the High Status group and that of the Low Status-plus-Control groups was also not statistically significant. The t was .40, p .34. Thus, there is no evidence here for an effect of the status variable (Prediction III).

In addition to Category H₁, two other types of hostile content were scored. Although no predictions were made concerning these categories, it was supposed that a differentiation of types of hostile content in the Rorschach protocol could be made corresponding to a differentiation of types in extra-test interpersonal behavior. If correlations which were

comparable to that between H_1 and the criterion were obtained between H_2 or H_3 and the criterion of overt aggression, serious doubt would be cast upon the conceptual validity of such differentiation.

Category H_3 represents all types of hostile content in which there is a connotation of hostile activity, but it is other than from testee to card. The correlation of this score with the criterion was not significantly different from zero, nor was the correlation of any of the subcategories of H_3 with group membership. In an over-all analysis of covariance of H_3 scores, none of the effects approached statistical significance, including the criterion group difference.

Category H_2 was scored for objects with a customary aggressive function. The rationale for scoring suggests that H_2 falls somewhere between the aggressive Rorschach response (H_1) and the kinds of hostility lumped together in H_3 . Perhaps it may be considered a potential or inhibited aggressiveness. The correlation between H_2 and the criterion fell short of statistical significance (ϕ was .19, p .07). However, in the over-all analysis of covariance of the H_2 score, the variance attributable to the criterion group difference was significant beyond the .01 level (F was 7.22, with df of 1 and 85). None of the other effects reached an acceptable level of statistical significance. For treatments, an F ratio of 1.89 was obtained; for interaction of treatments and the criterion group difference, the F ratio was 2.39.

It was hypothesized that testees examined under High Power conditions regard the test as having greater importance than testees in the Low Power and Control conditions. This assertion was tested by comparing the proportion of cases in the former group with the proportion in the latter groups reporting the examination as important when interviewed afterwards.

The number of Ss in the comparisons reported below was 78, because 18 cases either failed to respond to the question or responded too incoherently to permit scoring. The data are presented in Table 2.

The tendency to view the testing as important was positively correlated with being tested under the High Power condition (chi

TABLE 2
PROPORTION OF Ss REGARDING THE TEST AS "IMPORTANT" AS A FUNCTION OF CRITERION GROUP AND CONDITION OF EXAMINATION

Examination condition	Criterion Group		
	Assaultive	Non-assaultive	
High power	.69	.81	.75
Other	.62	.40	.53
	.64	.58	

square was 3.55, p .032). There is no evidence of a relation between criterion of aggression and the attribution of importance to the testing (chi square was .32; p was $> .50$).

Inspection of Table 2 suggested that there was an interaction effect of the marginal factors such that the differential attribution of importance to the testing is associated with the condition of testing in the NA but not in the A group. A test of this second order difference yielded a critical ratio of 1.91, which has a significance level of .056. This result suggests that the experimental conditions had an over-all effect on attitudes only because of a differential effect on the NA group.

DISCUSSION

The positive correlation between the H_1 scores and overt aggressive behavior lends support to the view that the Rorschach protocol is a sample of behavior in an interpersonal situation. The power and status variables failed to influence H_1 scores to a statistically significant degree.

In connection with this result, it is worth noting that the assumption was made, though only implicitly, that the aggressive behavior of the testees in extra-test interpersonal situations is influenced by the power and status of those around them. This has not been documented, and there is some reason to believe that for many of the testees, this relation does not obtain. That is, assaultive patients may be precisely those who are minimally influenced by these factors in their everyday behavior. The interview data also suggest that this factor operated to weight the experiment against the hypothesis.

The fact that the various categories of hos-

tility (H_1 , H_2 , and H_3) showed different relationships to the criterion suggests that the differentiation of types of hostility in the Rorschach is meaningful, and it provides a basis for clarifying the behavioral implications of content in Rorschach interpretation.

SUMMARY

The Rorschach was administered to 96 hospitalized psychotic veterans, 48 assaultive and 48 nonassaultive, under one of five conditions. Test records were scored for aggressive content (H_1) as well as two other types of hostile content. It was predicted that there is a positive correlation between the behavioral and test measures of aggression, and that the test measure is influenced by the experimentally varied power and status of the examiner. A significant positive correlation between the H_1 scores and the behavioral measure of aggression was found. The effect of power on the H_1 scores, while in the predicted direction, fell short of statistical significance. There is no evidence of an effect of examiner status on the H_1 score. Two other categories of hostile Rorschach response were studied: H_2 , scored for instruments with a customary aggressive function, correlated positively with the criterion (behavioral aggression). The other category of hostile response, H_3 , scored for responses connoting hostile activity, but not directed from testee to card, showed no relation either to the criterion or to the experimental variables. The differentiation of types of hostile content, based on an analogy to social behavior, seems to be a meaningful and potentially fruitful procedure.

REFERENCES

CALDEN, G., & COHEN, L. The relationship of ego-involvement and test definition to Rorschach test performance. *J. proj. Tech.*, 1953, 17, 300-311.
COCHRANE, W., & COX, GERTRUDE. *Experimental designs*. New York: Wiley, 1950.
Committee on Nomenclature and Statistics of the American Psychiatric Association. *Mental dis-*

orders. Washington: American Psychiatric Association Mental Hospital Service, 1952.
DOLLARD, J. Hostility and fear in social life. In T. M. Newcomb & E. L. Hartley (Eds.), *Readings in social psychology*. New York: Holt, 1947. Pp. 269-281.
EDWARDS, A. *Experimental design in psychological research*. New York: Rinehart, 1950.
ELIZUR, A. Content analysis of the Rorschach with regard to anxiety and hostility. *J. proj. Tech.*, 1949, 13, 247-284.
GIBBY, R. The stability of certain Rorschach variables under conditions of experimentally induced sets: I. The intellectual variables. *J. proj. Tech.*, 1951, 15, 3-26.
GORLOW, L., ZIMET, C., & FINE, H. The validity of anxiety and hostility Rorschach content scores among adolescents. *J. consult. Psychol.*, 1952, 16, 73-75.
JOEL, W. The interpersonal equation in projection methods. *Rorschach Res. Exch.*, 1949, 13, 479-482.
KIMBLE, G. A. Social influence on Rorschach records. *J. abnorm. soc. Psychol.*, 1945, 40, 89-93.
KLATSKIN, ETHELYN. Effect of the test situation upon the Rorschach record. *J. proj. Tech.*, 1952, 16, 193-199.
LORD, EDITH. Experimentally induced variations in Rorschach performance. *Psychol. Monogr.*, 1950, 64, No. 10 (Whole No. 316).
MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
MILLER, D. R. Prediction of behavior by means of the Rorschach test. *J. abnorm. soc. Psychol.*, 1953, 48, 367-375.
SANDERS, R., & CLEVELAND, S. E. The relationship between certain examiner personality variables and subjects' Rorschach scores. *J. proj. Tech.*, 1953, 17, 34-50.
SARASON, S. B. The test situation and the problem of prediction. *J. clin. Psychol.*, 1950, 4, 387-392.
SCHACHTEL, E. G. Subjective definitions of the Rorschach test situation and their effect on test performance. Contributions to an understanding of Rorschach's test. *Psychiatry*, 1945, 8, 419-448.
SINGER, J. L., MELTZOFF, J., & GOLDMAN, G. D. Rorschach movement responses following motor inhibition and hyperactivity. *J. consult. Psychol.*, 1952, 16, 359-364.
TOWBIN, A. P. Hostility in Rorschach content and overt aggressive behavior. Unpublished doctoral dissertation, Yale Univer., 1955.
WALKER, R. G. A comparison of clinical manifestations of hostility with Rorschach and MAPS test performance. *J. proj. Tech.*, 1951, 15, 444-460.

Received January 14, 1958.

THE GENERALITY OF SOCIAL PERCEPTION SCORES¹

MOHAMMED Y. QUERESHI

University of Illinois

RECENT studies of interpersonal perception (e.g., Bender & Hastorf, 1950; Bieri, 1953; Quereshi, 1957) claim to investigate such constructs as transference, empathy, real similarity, insight, etc., in terms of their operational equivalents. They have typically employed distance scores such as D^2 or D (Cronbach & Glaser, 1953) or simple absolute difference on the assumption that (a) the so-called operational measures of social perception are general over dimensions regardless of the content, and that (b) the sign of the difference score, i.e., the direction of the difference, is not important. A study bearing on the first of these assumptions was carried out by Cronbach, Hartman, and Ehart (1953), who report that "assumed similarity," indexed by absolute difference between two "cluster" scores, is a general attitude or mental set, independent of item content. Since they categorized items into clusters on the basis of logical analysis, however, their procedure could not insure independence of contents, and their claim that the "assumed similarity" measure was general over content-fields cannot be accepted as conclusive. In addition, few attempts have been made to estimate the reliability (internal consistency) of these scores with due regard to the content.

The present study was designed to break ground in this area by first analyzing the perceptual data into a number of orthogonal dimensions and then estimating (a) the generality and (b) the internal consistency reliability of a variety of measures of the perceptual relations (to be described in detail later). For each measure there were three main questions to be answered:

1. Is the measure reliable (consistent) within dimensions, i.e., consistent over items repre-

senting a particular content-field? If so, then to what extent?

2. Is the measure based on orthogonal factors general over various dimensions or content-fields? If so, then to what extent?

3. What is the effect of disregarding the sign of a difference or partial score on its generality over dimensions and its consistency within dimensions?

Answers to these questions have both methodological and practical significance for research in social perception. They should also throw light on conflicting results and confusion that exist in the present literature of this field.

METHOD

Subjects

During the initial stages of a program of research on therapeutic counseling at the Student Counseling Bureau of the University of Illinois, a number of tests, rating scales, and questionnaires were administered to persons who sought help from the Counseling Bureau. One of the rating scales administered to 204 students, 147 males and 57 females, was Ewing's Personal Rating Form, briefly described below. This sample consisted mostly of those who came to the Counseling Bureau for some sort of psychological help for their personal and emotional problems. However, some of the Ss came from an undergraduate psychology course, and we can assume that they were relatively well-adjusted.

Instruments and Data

The Ewing Personal Rating Form² consists of 100 adjectives such as "steady," "frank," "repressed," etc. (Quereshi, 1957). Each S was provided with three copies of this check list and was asked to rate himself, his father, and his mother on a five-point scale ranging from "Not Typical" to "Typical." The first part of the study was based on the self-ratings of the entire reference group of 204 Ss. Other-ratings (father and mother) were available for 98 males out of the total 204. The

² A 9-page document containing the table of rotated factor loadings, the Ewing Personal Rating Form, and keys for all the factors, has been deposited with the American Documentation Institute. Order Document No. 5809 from ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D.C., remitting \$1.25 for microfilm or \$1.25 for photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

¹ This paper is based on the author's master's thesis submitted to the Graduate College of the University of Illinois in 1957. Deep appreciation is due to Lee J. Cronbach for suggesting the problem and providing helpful suggestions and guidance throughout the course of this study. Thanks are also due to William Hays of the University of Michigan for a critical reading of the manuscript.

self-ratings and father-ratings of this smaller group formed the basis for the second part of the study.

Let the self-rating score on an item i for an individual a be designated as X_{ia} , and the father-rating score on the same item for the same individual as Y_{ia} . A variety of scores used in this study are based on or derived from these rating scores:

Factor score. The perceptual contents were to be analyzed first into orthogonal factors or dimensions. Each orthogonal factor was to be represented by a group of items. Thus a factor score for an individual a is equal to the sum of his rating scores on all the items defining a factor: These scores may be represented as

$$\sum_{i=1}^k X_{ia} \quad \text{or} \quad \sum_{i=1}^k Y_{ia}$$

depending upon whether we are using self-ratings or father-ratings to obtain the factor score. The respective factor scores may be designated C_{fza} and C_{fya} .

Factor difference score with regard to sign. This score is obtained by subtracting the factor score based on self-ratings from the factor score based on father-ratings, i.e.,

$$d_{ca} = \left(\sum_{i=1}^k Y_{ia} - \sum_{i=1}^k X_{ia} \right)$$

Factor difference score without regard to sign. This is obtained the same way as the immediately preceding score above, except that here the sign of the difference is disregarded. This score can be represented symbolically as

$$|d_{ca}|$$

Factor partial score with regard to sign. A factor partial score for an individual is based on the difference between the actual factor score based on father-ratings of that individual and the father-ratings factor score as predicted from the self-ratings factor score. Thus a factor partial score may be represented by:

$$P_{ca} = C_{fya} - \left\{ \bar{C}_{fya} + r_{C_{fya}C_{fza}} \frac{\sigma_{C_{fya}}}{\sigma_{C_{fza}}} (C_{fza} - \bar{C}_{fza}) \right\}$$

Factor partial score without regard to sign. This is obtained in the same way as the immediately preceding score, except that in this case the sign of the partial score is disregarded, i.e.,

$$|P_{ca}|$$

Factor D^2 and D score. The factor D^2 for an individual is obtained by squaring the difference between father-ratings and self-ratings for that individual and summing them over items comprising a factor. Thus, the D score for an individual a for factor f is

$$D_{fa} = \sqrt{\sum_{i=1}^k (X_{ia} - Y_{ia})^2}$$

Procedure

The first part of this study consisted of factor analysis of covariances based on self-ratings of the 204 Ss.

Each item was treated as a variable, and covariances were computed for the self-ratings. The matrix of covariances was then analyzed, so as to extract orthogonal dimensions by means of Thurstone's Centroid Method. As many of the factors as possible were retained in order to account for the maximum possible covariance. Out of 10 unrotated factors, 7 were retained and subjected to quartimax rotation in order to approximate simple structure. Examination showed that quartimax rotation did not approach the desired degree of psychological meaningfulness in the case of Factors V and VII, which were therefore rotated graphically to improve meaningfulness.

Inspection of the item factor loadings after rotation showed that a number of them were extremely low. Some of the items could thus be excluded from the later operations without incurring the loss of any substantial covariance. Factor keys were therefore composed from positive or negative loadings, i.e., loadings $\pm .50$ or higher on Factor I and $\pm .30$ or higher on Factors II through VII. These keys formed the basis for the factor scores on self- and father-ratings for the 98 male Ss. The items thus included in the factor keys³ gave unit positive or negative weight to each raw score, since an empirical comparison by Trites and Sells (1955) between this unit weighting method and fractional weighting concluded that little could be gained by resorting to the latter.

In order to obtain estimates of the internal consistency, each factor key was subdivided by sorting the selected items into two equated groups, and assigning items to half-factor keys (A and B) in such a manner that they were fairly balanced with respect to number of items and magnitude and sign of loading.

Both the full-factor scores and the half-factor scores (based on items in either key A or B of a factor) were then computed for all the factors, and the pairwise correlations among the 14 half-factor scores determined, both for the self-ratings and for the father-ratings. Half- and full-factor difference and partial scores were next computed, both with and without regard to sign.

The Item D scores were also computed for both the half-factor keys and the full-factor keys. This procedure yielded 14 half-factor Item D scores and 7 full-factor Item D scores.

Correlations were computed among half-factor as well as full-factor scores of all the types mentioned above. These provided evidence both on the reliability (internal consistency) of these scores for various content-fields or dimensions and their generality across dimensions.

RESULTS AND DISCUSSION

As already reported, the item covariances were factor analyzed using Thurstone's centroid method and quartimax rotation. Seven

³ The number of items in each factor key is as follows: Factor I = 24; Factor II = 24; Factor III = 25; Factor IV = 23; Factor V = 17; Factor VI = 17; Factor VII = 11.

TABLE 1

SPLIT-HALF RELIABILITIES, MEANS AND STANDARD DEVIATIONS OF VARIOUS SCORES USED IN THIS STUDY
(*N* = 98)

Scores	Factor I (Unhappy)				Factor II (Outgoing)				Factor III (Contrary)				Factor IV (Hardworker)			
	<i>r</i>	<i>r</i> *	<i>M</i>	<i>SD</i>	<i>r</i>	<i>r</i> *	<i>M</i>	<i>SD</i>	<i>r</i>	<i>r</i> *	<i>M</i>	<i>SD</i>	<i>r</i>	<i>r</i> *	<i>M</i>	<i>SD</i>
Scores based on self-ratings	.89	.94	63.5	15.4	.65	.79	31.6	12.2	.84	.91	42.6	13.6	.75	.86	52.7	11.4
Scores based on father-ratings	.87	.93	45.8	16.3	.56	.72	31.7	11.3	.85	.92	47.2	17.3	.85	.92	56.2	14.7
Difference scores with regard to sign	.81	.90	-17.7	17.3	.55	.71	00.0	15.3	.80	.89	4.6	19.8	.83	.91	3.5	17.7
Absolute difference scores	.74	—	20.2	14.3	.38	—	11.9	9.7	.57	—	15.9	12.7	.70	—	14.1	11.2
Partial scores with regard to sign	.84	—	00.0	14.9	.55	—	00.0	11.3	.83	—	00.0	17.0	.86	—	00.0	14.6
Absolute partial scores	.64	—	12.1	8.7	.37	—	8.7	7.1	.69	—	13.8	9.9	.74	—	11.3	9.2
<i>D</i> scores	.74	—	6.7	4.2	.59	—	6.2	3.5	.63	—	6.6	4.1	.78	—	5.7	3.9

Scores	Factor V (Conventional)				Factor VI (Cyclothyme)				Factor VII (Creative)			
	<i>r</i>	<i>r</i> *	<i>M</i>	<i>SD</i>	<i>r</i>	<i>r</i> *	<i>M</i>	<i>SD</i>	<i>r</i>	<i>r</i> *	<i>M</i>	<i>SD</i>
Scores based on self-ratings	.45	.62	22.2	7.1	.70	.82	-15.9	8.9	.47	.64	22.7	5.3
Scores based on father-ratings	.42	.59	22.5	7.4	.75	.86	-14.2	10.8	.57	.73	19.8	5.4
Difference scores with regard to sign	.49	.66	00.3	10.8	.73	.84	1.8	12.5	.45	.62	-2.9	6.0
Absolute difference scores	.03	—	8.4	6.8	.49	—	10.0	7.7	.24	—	5.1	4.3
Partial scores with regard to sign	.40	—	00.0	7.3	.76	—	00.0	10.6	.53	—	00.0	5.0
Absolute partial scores	.02	—	6.0	4.3	.54	—	8.5	6.3	.09	—	4.2	2.6
<i>D</i> scores	.49	—	4.2	2.7	.55	—	4.0	2.4	.30	—	2.4	1.5

* Corrected by Spearman-Brown Formula.

factors that were selected for later investigation have been named as follows:

Factor I	Unhappy
Factor II	Outgoing
Factor III	Contrary
Factor IV	Hard worker
Factor V	Conventional
Factor VI	Cyclothyme
Factor VII	Creative

It was apparent from the beginning that the factor scores thus obtained would not be orthogonal to one another, for one reason because some items were included in more than two factor keys. Such a case is illustrated by Factors III and VI. Since these two factors are highly correlated with one another, Factor VI may either be dropped or combined with Factor III in subsequent research without losing any significant information. On the whole, as the figures for Scores 1 and 2 indicate, the seven factors have a median absolute intercorrelation of approximately .30. Orthogonality of factor scores could only have been

achieved in this study by the laborious assignment of precise item weights.

Reliability (Internal Consistency) of Various Scores

Table 1 presents the split-half reliabilities, means, and standard deviations of the various scores. For the first three types of scores, reliability has been corrected by the Spearman-Brown formula and is reported together with the uncorrected estimates. For the last four scores, the use of the Spearman-Brown formula is questionable because these scores do not meet its assumptions.⁴

The reliability estimates (Table 1) for factor scores based on self-ratings and for factor scores based on father-ratings are approxi-

⁴ For example, in the case of *D* score, the correlation between two parallel keys of double length would be

$$r \sqrt{D_{fa}^2 + D_{fa}^2} \sqrt{D_{fa1}^2 + D_{fa1}^2}$$

whereas the S-B equation estimates

$$r(D_{fa} + D_{fa})(D_{fa1} + D_{fa1})$$

TABLE 2
MEDIAN AND MEAN ABSOLUTE INTERCORRELATIONS OF VARIOUS SCORES
(Estimates of factor-intercorrelations and generality of scores)

Scores	Factor I (Unhappy) with the other 6 factors		Factor II (Outgoing) with the other 6 factors		Factor III (Contrary) with the other 6 factors		Factor IV (Hard worker) with the other 6 factors	
	Median	Mean	Median	Mean	Median	Mean	Median	Mean
Factor scores based on self-ratings	.35	.27	.36	.32	.21	.27	.13	.19
Factor scores based on father-ratings	.36	.37	.32	.35	.34	.42	.37	.36
Factor difference scores with regard to sign	.31	.25	.36	.36	.26	.36	.28	.29
Factor absolute difference scores	.09	.12	.19	.21	.18	.25	.11	.12
Factor partial scores with regard to sign	.37	.33	.34	.36	.35	.42	.39	.37
Factor absolute partial scores	.13	.19	.20	.20	.24	.29	.11	.15
Factor <i>D</i> scores	.56	.56	.72	.72	.52	.63	.66	.64

Scores	Factor V (Conventional) with the other 6 factors		Factor VI (Cyclothyme) with the other 6 factors		Factor VII (Creative) with the other 6 factors	
	Median	Mean	Median	Mean	Median	Mean
Factor scores based on self-ratings	.27	.29	.25	.31	.09	.17
Factor scores based on father-ratings	.34	.34	.35	.45	.27	.29
Factor difference scores with regard to sign	.35	.31	.29	.39	.24	.23
Factor absolute difference scores	.34	.29	.14	.25	.20	.19
Factor partial scores with regard to sign	.33	.33	.38	.44	.23	.28
Factor absolute partial scores	.25	.22	.17	.27	.08	.12
Factor <i>D</i> scores	.70	.68	.63	.63	.63	.62

mately the same.⁵ Also, the reliability estimates for factor difference and partial scores with regard to sign are strikingly close to one another and almost identical with those for factor scores based on self- and father-ratings, respectively. The corrected reliability estimates for the first three types of scores listed in the table, for four out of seven factors, are in the vicinity of .90. For the remaining three factors (II, V, and VII) they range between .59 and .79.

It is significant that the uncorrected reliability estimates for absolute difference and partial scores for all factors are consistently lower than the uncorrected reliability estimates for the same scores when the sign is considered. In every case, the reliability esti-

mates for the former scores are lower than those for the latter, and in some cases, e.g., Factors V and VII, the reductions in magnitudes are significant at .05 or less. Ignoring the sign of the difference or partial score thus results in loss of information.

The split-half reliability estimates for *D* scores are generally higher than those for the absolute difference and partial scores, but are consistently lower than those for the other four types of measures for all factors except II and V. For these two factors, the reliabilities are either close to those for these four measures or a little higher. On the whole, the tendency is toward a decrease in the magnitude of reliability for *D* scores when compared with the four types of scores referred to above, though further investigation is needed on this point.

Generality of Various Scores

⁵ It may be pointed out that the reliability of the difference scores with regard to sign could be computed from the reliability of full-factor scores based on self-ratings and of those based on father-ratings by the simple use of formulas available for this purpose. The values thus obtained would agree with those we have calculated directly.

Since the basic factor scores are intercorrelated, the derived scores based on them should also be intercorrelated by statistical necessity. The results in Table 2 show that the

TABLE 3
INTERCORRELATIONS OF SCORES FOR PAIRS OF ORTHOGONAL FACTORS

Factor Pair	Intercorrelations of Scores*						
	<i>s</i>	<i>f</i>	<i>d</i>	<i>/d/</i>	<i>P</i>	<i>/P/</i>	<i>D</i>
I with VII	-.08	-.17	-.03	.10	-.12	.10	.61
III with VII	-.02	-.18	-.19	.20	.01	.09	.52
I with V	.08	-.02	-.28	.09	-.21	.01	.49
IV with VI	.11	-.30	-.22	-.07	-.31	-.00	.55
IV with V	-.06	-.21	-.09	.12	-.20	.24	.72
III with IV	.14	-.16	-.15	.02	-.18	.11	.52
II with VI	.17	.09	.23	.08	.13	-.08	.71
VI with VII	-.09	-.27	-.13	.20	-.21	.05	.54

* *s* stands for correlation based on self-ratings; *f* for correlation based on father-ratings; *d* for difference score with regard to sign; */d/* for difference score without regard to sign; *P* for partial score with regard to sign; */P/* for partial score without regard to sign; and *D* for *D* score as explained elsewhere.

Note.—The above correlations are based on 98 cases, and *p* for an *r* of .20 = .05.

median absolute correlations representing the generality of factor difference and of factor partial scores (both with and without regard to sign) are hardly different from those of the factor scores based on self-ratings and father-ratings. Since the last two scores indicate the degree to which the factors are intercorrelated, the factor difference and factor partial scores would appear to be general only to the extent to which the perceptual contents overlap.

The *D* scores, unlike the factor difference and partial scores with regard to sign, show appreciable generality over dimensions. Even after making due allowance for the intercorrelations of factors on which *D* scores are based, the *D* scores are significantly general over dimensions. The difference between the median of median *D* score absolute interfactor correlations (.63) and the median of median absolute interfactor correlations based on factor scores for self- and father-ratings (.33) is statistically significant at the .05 level. In the light of the available evidence, the *D* score can thus be justifiably used in research on social perception without regard to the content dimensions involved.

One of the initial questions concerned the effect of ignoring the sign of a difference or partial score on its generality. A comparison of the median *r*'s for factor difference and partial scores with regard to the sign with the median *r*'s of the absolute difference and partial scores indicates a consistent and appreciable drop in generality for all the factors. In the case of partial scores, for some factors the drop in generality estimates (from .37 to .13 for Factor I, and from .39 to .11 for Factor IV,

respectively) is significant at .05 level (*P* of a difference of .24 between two correlated *r*'s being $\leq .05$). These results definitely lend support to the view that the sign of a difference or partial score is important for its generality and that ignoring it results in the loss of relevant information.

The drop in generality estimates of difference and partial scores, when the sign is disregarded, is also apparent in Table 3, which shows the intercorrelations of scores for pairs of orthogonal factor scores. Here the correlations are reported with their sign, unlike Table 2 which showed medians of absolute intercorrelations of factors. The estimates of correlations for most pairs are appreciably reduced when the sign of the score is disregarded. This tendency, however, is not as sharply illustrated in these results as in those of Table 2. The *D* score for these pairs of factors is significantly general, the *r*'s ranging between .52 and .72.

An Evaluation

Since the method of analyzing social perception scores adopted in this study has been tried here for the first time, it would be premature to assert that its usefulness for research in this field has been demonstrated. Moreover, the approximation of factor scores by factor keys employing unit weights is also a preliminary technique subject to modification and refinement. Caution is in order in interpreting the results concerning the reliability and generality of various scores used in research on interpersonal perception, and concerning the effect of ignoring the direction of difference on their reliability and generality. However, it

can be justifiably stated that the evidence so far available does not support the generality of most of these scores, and indicates that the sign of difference or partial measures can be ignored only with the loss of relevant information.

The generality of *D* scores, as supported by the present investigation, has important bearing on their use in social perception research. These results lend support to the view that the *D* score can be used without resorting to dimensional analysis.

SUMMARY

This study of self-ratings and other-ratings was designed to investigate (a) whether it is justified to assume the generality of various distance scores regardless of the content involved, (b) the reliability of these scores, and (c) the effect on their generality and reliability when the sign of any of the distance measures is ignored. The method of data analysis employed is comparatively new and has been tried here for the first time.

The results indicate that what generality most of these scores possess can be attributed to overlap between the content areas on which the ratings are based. Except for one of the measures (*D* score) used in this study, all follow the same pattern. Appreciable drop in generality is consistently found when the sign of difference is disregarded. The generality of *D* score seems to justify its use as a defensible global measure.

Reliability estimates are also appreciably

reduced in magnitude when the signs of the distance measures are not taken into consideration (the *D* score is no exception here). The results do not approach desired statistical significance in all cases, but the trends are clearly and consistently indicative of the loss of relevant information when the sign of the measure is disregarded.

REFERENCES

- BENDER, J. E., & HASTORF, A. H. The perception of persons forecasting another person's responses on three personality scales. *J. abnorm. soc. Psychol.*, 1950, **45**, 556-561.
- BIERI, J. Changes in interpersonal perception following social interactions. *J. abnorm. soc. Psychol.*, 1953, **48**, 61-66.
- CRONBACH, L. J. Proposals leading to analytic treatment of social perception scores. In R. Tagiuri & L. Petrullo (Eds.), *Person perception and interpersonal behavior*. Stanford, Calif.: Stanford Univ. Press, 1958. Pp. 353-379.
- CRONBACH, L. J. Processes affecting scores on "understanding of others" and "assumed similarity." *Psychol. Bull.*, 1955, **52**, 177-193.
- CRONBACH, L. J., & GLASER, GOLDINE C. Assessing similarity between profiles. *Psychol. Bull.*, 1953, **50**, 456-474.
- CRONBACH, L. J., HARTMAN, W., & EHART, MARY E. Investigation of the character and properties of assumed similarity measures. Bur. of Educ. Res., Univer. of Illinois, 1953. (Mimeo.)
- QUERESHI, M. Y. An empirical study of the generality of social perception scores. Unpublished master's thesis, Univer. of Illinois, 1957.
- TRITES, D. K., & SELLS, S. B. A note on alternative methods for estimating factor scores. *J. appl. Psychol.*, 1955, **39**, 455-456.

Received December 30, 1957.

SUBJECTIVE CERTAINTY AND RESISTANCE TO CHANGE¹

DONALD S. BOOMER

National Institute of Mental Health

THIS study is concerned with the psychological bases for resisting or conforming to group influence. The point of departure was the desire to tap subjects' "private" experience during the application of group pressure in order to learn more about the psychological processes involved.

Experimental approaches to conformity most typically identify conforming subjects by some simple behavioral criterion, e.g., revising a discrepant estimate or changing a vote to agree with a group consensus. Dichotomizing subjects (Ss) thus according to their behavior at a choice point lends itself readily to the assumption that the underlying psychological processes are similarly dichotomized. Such an assumption ignores the fact that the overt behavior being noted is the only observable aspect of a complex psychological process, and identical behavior among subjects does not necessarily reflect a corresponding identity, or even similarity, in the psychological processes involved.

This principle has been cogently demonstrated by Asch (1952) in investigating the problem of the "minority of one vs. a unanimous majority." His basic experimental design categorized Ss behaviorally as "independents" and "yielders." Postexperimental interviews, however, revealed among his Ss three qualitatively different kinds of processes that led to yielding and two that led to maintaining independence. These differences had not been discernible in the behavioral phase of the experiment.

Our own study employed periodic estimates of certainty² by the S during the period of experimental group pressure in order to see what differences they might reveal among people whose observable behavior was similar.

¹ This research is a portion of a Ph.D. thesis submitted to the Department of Psychology, University of California. The author is indebted to the members of his doctoral committee, Read D. Tuddenham, Richard S. Crutchfield, and Gordon W. Hearn, for their guidance and criticism.

² The writer is indebted to Jack Block for suggesting this specific measure.

The procedure was adapted from Crutchfield (1951) who in turn had adapted it from an unpublished problem-solving task devised by Bavelas. A brief summary of Crutchfield's procedure provides the necessary context for a statement of the central issue of the present study.

His Ss worked in groups of six on a group problem that involved fitting together, jigsaw-fashion, irregularly shaped cardboard pieces into squares. Each S had one such square to make; the task required each S to complete his square from the limited set of cardboard pieces available in common to the group. The Ss were told that they would be exchanging pieces with one another in order to complete their squares.

This process of exchange was accomplished indirectly through the experimenter, enabling him to maintain strict control of the situation. Actually no real transactions were permitted; the experimenter posed requests and made pieces available for exchange according to a preset schedule, maintaining the illusion for the Ss that they were dealing with one another.

Each S was permitted to complete his square early in the experiment, but was led to believe that others were not yet finished, since the experimenter continued to make rounds of the group presenting simulated "requests" for pieces still needed. After a few such circuits of the group, the experimenter presented a request for one of the pieces that each S was using in his completed square. The crucial determination was how many successive requests were required of a S before he broke his square to yield the piece requested.

Some Ss "broke" at the first such request; at the other extreme one S broke only after 20 successive requests. Crutchfield (1951) grouped his Ss into three categories: fast, medium, and slow breakers. In his report he says, "The test confronts each person in the group with an identical problem such that he is placed in a conflict arising from competing demands of cooperation, of independent analysis of the problem, and of maintenance of a

stable configuration. The person's resulting behavior in the conflict situation is objectively and simply scored" (pp. 587 f.). The inferential reconstruction of the psychological situation is then temptingly straightforward. The first request for the crucial piece generates this conflict; each *S* is torn between the desire to cooperate and the desire to maintain his stable configuration. Whichever of these motives is the more potent determines whether he will break early or late.³

Consideration of Asch's interview findings and James' Psychologists' Fallacy raises questions about this interpretation. Part of the variability among these *Ss* might be due to their relative readiness to become aware that the situation posed a conflict for them. Some slow breakers might be slow not because the stability of their own solutions was more important to them than cooperation, but because they had protected themselves from experiencing this conflict by their perceptual organization of the situation. Data were sought concerning the *Ss*' periodic assessments of their own feeling of confidence during the experiment in the hope that they would indicate when and if *Ss* were experiencing conflict. The larger study (Boomer, 1954) on which this report is based was also concerned with some other parameters of Crutchfield's experimental situation which will not be reported here.

METHOD

Subjects

The *Ss* were 72 undergraduate college students, 36 men and 36 women, drawn from a large lower division course in personal and social adjustment. They were not volunteers; all *Ss* were required to participate in one experiment as a part of their course work. These *Ss* were randomly divided into 12 experimental groups: six groups of six men and six groups of six women.

Procedure

The experimental procedure was modelled quite closely on the Crutchfield (1951) "broken

³ In the interests of clarity, the third factor, "independent analysis of the problem," is not considered, since the data on estimation of certainty cannot be brought to bear upon this factor as a possible aspect of the conflict.

square" procedure described above, with the modification that numbers were substituted for jigsaw pieces as the experimental materials. Relevant parts of the instructions, adapted directly from Crutchfield's report, are reproduced here in order to explain the procedure and to give the flavor of the situation. These instructions were read aloud to the experimental groups of six *Ss* each:

This is a group problem-solving situation with a group goal. There are no individual goals in it separate from the group goal. The six members of this group will work on a common group problem, but the only communication among you will be via the experimenter, who will circulate between you throughout the task, serving as a messenger.

The materials are a set of small numbered wooden discs. Each of you will begin with five randomly selected numbers, and the group task is for each of you to acquire by trading back and forth either a sequence of five numbers—5-6-7-8-9, say, or a set of five numbers, five 3's, say. Each person's part of the total group problem is to assemble one such set or sequence. The group goal is considered achieved only at that point when *all six* of you have a set or a sequence.

The available numbers range from 1 to 15, but you have no way of knowing how many of each there are. Actually the numbers have been so chosen that there are many ways for one or more of you to acquire a set or a sequence, but very few ways in which all of you can get simultaneous solutions. This is the real essence of the group task.

To make this point clear I will show you a similar set-up in which the task is to acquire a set or a sequence of only three numbers. (The following diagram is drawn on the blackboard:)

- | | |
|----------|----------|
| A. 1 2 3 | D. 6 6 6 |
| B. 2 3 4 | E. 7 7 7 |
| C. 3 4 5 | F. 9 9 9 |

These six people, represented here by the letters A, B, C, etc., have traded their original numbers back and forth until each of them has a solution. This is one of the few ways in which this set of numbers could be arranged to provide six simultaneous solutions. If someone had chosen to collect 3's, however, it would have been impossible for A and B to get sequences. Or had someone chosen to save this sequence, 5-6-7, it would have been impossible for D and E to get sets.

All exchanging of numbers took place via the experimenter (*E*), who carried a small tray from *S* to *S*. One section of the tray held numbers that were available for exchange; another section of the tray showed numbers that had been purportedly requested by members of the group. This provision for indirect communication through *E* permitted the deception on which the experiment depended.

Ss were seated in a semicircle, facing outward

so they could not see one another directly nor *E*, who walked behind them circulating the tray. *E* was thus free to manipulate the numbers on the tray in such a way that a standard stimulus situation could be maintained for all *S*s throughout the experiment. On each round, or circuit of the group, *S*s all faced an identical exchange-request situation. When an *S* exchanged a piece or requested one, *E* secretly restored the standard situation before presenting the tray to the next *S*. Thus, although each *S* believed he was freely communicating with the other members of the group he was actually dealing only with *E* in a controlled situation.

The sequence of events was this: All *S*s began the task with identical "hands," 4-5-7-7-10, which they were led to believe was a chance draw. By offering a 6 at the first round and an 8 at the third, *E* led each *S* to build the sequence, 4-5-6-7-8. Rounds 4 and 5 offered and requested numbers above 10, which did not concern any *S*.⁴

At Round 6 each *S* was confronted with a request, ostensibly from a teammate, for a 7. Those *S*s who broke their sequence to yield the 7 were permitted to build a new sequence, 2-3-4-5-6, during the next two rounds, while the pressure of the critical request was maintained on each round for the others. In the event that some *S*s in the group still had not broken their sequences after 15 crucial rounds, the experiment was discontinued, and the group was told that they had succeeded in the group task. This was credible since each *S* had a solution, either his original 4-8 sequence or the replacement 2-6 sequence.

Each *S* was also furnished at the outset with a pad of small protocol slips numbered to correspond with the rounds. He filled out the appropriately numbered slip after every round indicating what trade he had made, if any, and what number he had requested, if any. If he had neither traded nor requested, he checked "pass" on the slip. On each round of the group, *E* collected the slips from the previous round, checking them for accuracy. These provided a

succinct record of each *S*'s transactions on each round.

In addition to recording his transactions, *S* estimated his feeling of confidence⁵ at each round. At the bottom of each protocol slip was a three-inch line with 0 at the left, 50 in the center and 100 at the right. Along this line *S* was asked to write a number representing his degree of certainty that his transaction on this round had been "the best thing to do" for him and his group. These certainty estimates will hereinafter be referred to as CEs.

A final source of information was the free comments which *S*s were urged to write on the backs of these slips. Since *E* got around the group only once every 3 or 4 minutes, *S*s had ample time to formulate their impressions. These written comments were not analyzed as data, but did yield some useful impressions about *S*s' perception of the unfolding situation.

RESULTS

The number of the trial at which *S* yielded to the request for a 7 will be termed his "break score." Break scores ranged from 6, the first request round, to 20, the last round.⁶ Twenty-three *S*s had not yielded at the last round; their scores were recorded as NB (nonbreak).

The distribution thus obtained, like Crutchfield's, was markedly skewed. The following classification of scores was employed in the statistical analysis:

1. "Early breakers," $N = 17$, *S*s yielding on Rounds 6 and 7, the first two request rounds
2. "Moderate breakers," $N = 17$, *S*s yielding on rounds 8 and 9
3. "Late breakers," $N = 13$, *S*s yielding on rounds 10 through 20
4. "Nonbreakers," $N = 23$.

Certainty estimates. The mean CEs of nonbreakers and breakers are generally discrepant throughout the experiment, most markedly at certain critical junctures. Table 1 shows the mean CEs for these two groups over the first six rounds. Beginning at round three the mean

⁵ Although these determinations were originally termed "confidence estimates," they are called "certainty estimates" in this report to avoid possible confusion with the statistical concept.

⁶ Two *S*s unaccountably broke their sequences on Round 5, just before the crucial rounds began. Since their subsequent experience was not comparable to that of the other *S*s, their data were discarded, reducing the N to 70.

⁴ These two noncrucial rounds were deliberately inserted to provide for the possibility that some *S*s might not accept the forced sequence on Rounds 2 and 3. Only three *S*s failed to complete the sequence at Round 3, and they completed it on Round 4.

TABLE 1

MEAN CERTAINTY ESTIMATES OF BREAKERS AND NONBREAKERS FOR ROUNDS 1 THROUGH 6

Round	Breakers		Nonbreakers		<i>t</i>	<i>p</i>
	<i>N</i> ^a	\bar{X}	<i>N</i> ^a	\bar{X}		
1	47	59.0	22	70.8	1.76	NS
2	47	47.8	22	60.4	1.96	NS
3	46	73.6	21	96.8	5.57	.001
4	45	77.7	21	92.2	2.96	.01
5	42	78.4	21	91.4	2.65	.02
6	36 ^b	72.0	22	86.6	2.45	.02

^a The discrepancies in the *N*s for Rounds 1-5 are due to *S*s' occasional failure to record CEs on their protocols.

^b CEs of the 10 *S*s who broke on this round are not included in this mean.

TABLE 2

FREQUENCY OF SUBJECTS IN FOUR SCORE CATEGORIES EXPRESSING CERTAINTY ESTIMATES OF 100 AT ROUND 3

Score Category	Number of Subjects Expressing Certainty Estimates of 100	Number of Subjects Expressing Certainty Estimates of < 100	Totals
Early breakers	4	12	16
Moderate breakers	5	12	17
Late breakers	6	7	13
Nonbreakers	19	3	22
Totals	34	34	68 ^a

Note.—Chi square = 19.28; $p < .001$.

^a Two *S*s neglected to record any certainty estimate on this round.

CEs for breakers and nonbreakers are significantly different.

Contributing heavily to their higher mean CEs was the nonbreakers' more frequent use of 100 as an estimate. This finding suggested another approach to these data. The difference between 100 and any smaller number, even 99, represents a distinction between the expression of absolute certainty and the admission of some uncertainty. Let us consider the relative incidence of CEs of 100 in the various break-score categories.

Two junctures are presented as being particularly salient in reflecting differential response to the experimental situation. The first is Round 3, when *S* completed his first sequence. Table 2 presents the frequency of 100% CEs on this round for each of the four break-score categories. The frequency differences among the categories yield a chi square which is significant at the .001 level of confi-

TABLE 3

FREQUENCY OF SUBJECTS IN FOUR SCORE CATEGORIES EXPRESSING CERTAINTY ESTIMATES OF 100 AT ROUND 6

Score Category	Number of Subjects Expressing Certainty Estimates of 100	Number of Subjects Expressing Certainty Estimates of < 100	Totals
Early breakers	0	6	6 ^a
Moderate breakers	5	12	17
Late breakers	5	8	13
Nonbreakers	13	9	22
Totals	23	35	58 ^b

Note.—Chi square = 8.17; $p < .05$.

^a CEs of the 10 *S*s who broke on this round are not included in this table.

^b Two *S*s neglected to record any certainty estimate on this round.

dence. It is clear that whether or not *S* expressed complete confidence at this very early point is closely related to whether or not he will ultimately yield.

The other juncture selected for scrutiny was Round 6, when *S* first encountered the request for the 7 he was using in his sequence. The CE on this round reflects *S*'s initial reaction to this, the crucial conflict posed by the experimental situation. Ten *S*s yielded to this first request; their CEs are excluded from consideration here, since their actual situation was changed by their having broken. The remaining 35 breakers behaved like the nonbreakers on this round in that they maintained their solutions despite this initial request, but they expressed less certainty about it than did the nonbreakers. The relative frequencies of CEs of 100 for the four groups are shown in Table 3. The frequency difference among these groups yields a chi square which is significant at the .05 level of confidence.

DISCUSSION

To summarize the CE data, they demonstrate that *S*s who express perfect certainty at Round 3, upon completing the initial sequence, are statistically less likely to break this sequence later than are those who express some uncertainty at this juncture. Similarly, at Round 6 those *S*s expressing 100% certainty when declining to yield to this first request for the 7 are less likely to yield on any successive

round than are those who express some uncertainty.

If all *Ss* were indeed experiencing the presumed conflict between the demands of cooperation and of maintaining a stable configuration there would be no reason to expect systematic CE differences between breakers and nonbreakers. If the demand for the 7 at Round 6 in fact poses a dilemma for all *Ss*, this should be reflected in lowered CEs among nonbreakers as well as breakers. Instead the data show that some *Ss* express no uncertainty about maintaining their solutions intact in the face of this request and that these *Ss* are predominantly nonbreakers.

The instructions, it will be recalled, do not explicitly anticipate this situation in which *S* is asked for a number already imbedded in his completed sequence. When *S* is confronted with this situation at Round 6, its impact on him depends on his perceptual organization of the total situation, including his recall of the instructions and the meaning he places on his own completed sequence.

Differences among *Ss* with regard to the meaning of having achieved a sequence are reflected in their CEs at Round 3. The expression of perfect certainty at this point is unwarranted, considering *all* the information which was provided in the instructions. It had been clearly stated that the task was a group task, that many individual solutions were possible but very few simultaneous group solutions. It was even explicitly demonstrated how one premature individual solution could block the group solution.

If *S* fails to recall this critical aspect of the task he can view his personal solution as the completion of his part of the group task and register 100% certainty that completing this sequence was the best thing to do for himself and his group. With such a set he can, without discomfort, interpret the subsequent request for a 7 as not directed to him, and he can keep his solution intact, still with 100% certainty that this is the best course. Such *Ss* are not in conflict between the competing demands of cooperation and of maintaining a stable configuration; they have perceptually organized the situation in such a way as to eliminate conflict.

Support for this interpretation of the psycho-

logical implications of high and low CEs was provided by the free written comments of some of the more verbal *Ss*. Following are clear statements by four *Ss* at Round 3, when each had completed his sequence:

Breaker, (CE 30): "Have completed a run from 4-8. I will see what numbers are available and what ones are requested before planning to keep the run or not."

Breaker, (CE 80): "I now have the sequence 45678 I hope that it does not interfere with any other's and have anxiety in thinking what I will do if someone needs one of mine."

Nonbreaker, (CE 100): "My part of the experiment is complete therefore I cannot trade."

Nonbreaker, (CE 100): "Finished my accomplishment. Satisfaction."

A similar contrast is provided by a selection of comments at Round 6, the first request round. None of these four *Ss* had yielded to this initial request, but the comments and the CEs of those who were to break later indicate their awareness of the conflict.

Breaker, (CE 40): "Had number needed but giving up this would throw my sequence entirely—if, however, it is requested again I will prob. give it up."

Breaker, (CE 50): "7 has been requested don't know whether I should give it up and try for a new sequence."

Nonbreaker, (CE 100): "I couldn't help request section without ruining my sequence. Thus 100 and reason I passed."

Nonbreaker, (CE 100): "I hope the other boys are doing as well as I."

To return to the central problem of the study: what is the nature of the psychological process underlying yielding or not yielding to group pressure? The certainty estimate data lend support to the hypothesis that an important part of the behavioral variability in this situation can be attributed to *Ss*' differential perceptual organization of the situation. The previous interpretation based on the response to competing demands must be modified to take account of *Ss*' relative propensity to organize complex situations in such a way as to minimize the internal conflict.

SUMMARY

This study examined the relationship between *Ss*' expressed level of certainty during an experiment and their observable behavior. Seventy-two undergraduates, randomly as-

signed to groups of six, were subjected to a quasi group-interaction situation.

Ss were led to believe that they were exchanging numbered counters with one another, through a messenger, to the end that all six members could achieve simultaneous solutions to a problem. Actually they were dealing only with *E* who throughout the task maintained a standard pattern of offers and requests for all. The crucial determination was whether or not *S* would break his own solution to yield to a simulated request from another group member for one of the numbers he was using.

At each transaction with the "messenger," Ss were required to estimate their level of certainty on a scale from 0 to 100. Certainty estimates at selected early junctures were then examined in the light of whether or not the Ss had ultimately yielded.

At the point when all Ss achieved their in-

dividual solutions and before any request had been made of them, some Ss expressed 100% certainty. These Ss later proved to be less likely to yield to the critical request than were those Ss who had expressed some uncertainty. This evidence is used to support an interpretation of yielding to group pressure which takes account of Ss' perceptual modes of organizing complex, ambiguous situations.

REFERENCES

- ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
- BOOMER, D. S. Some determinants of the response to a conflict between personal and group goals: A methodological investigation of the quasi-group. Unpublished doctoral dissertation, Univer. of Calif., 1954.
- CRUTCHFIELD, R. Assessment of persons through a quasi group-interaction technique. *J. abnorm soc. Psychol.*, 1951, **46**, 577-588.

Received January 8, 1958.

THE EFFECTS ON OPINIONS OF A CHANGE IN SCALE JUDGMENTS¹

WALTER WEISS

Boston University

A PERSON'S position on a particular issue is manifested in part by the distribution of opinion items that he accepts. If the items are ordered on the issue dimension, his range of acceptance comprises an ordered portion of the total continuum. That is, the items with which one person agrees may range from extremely pro to mildly pro, while a second may accept items from extremely pro to neutral, and a third, items from mildly pro to mildly con, etc. Persons' ranges of acceptance may thus differ with respect not only to extent, but also to the location on the continuum of the pro and con limits beyond which items are not accepted. Obviously, many factors may influence the location and extent of a person's range of acceptance. As long as coercion is not involved, however, the selection of acceptable items should depend *in part* on the person's judgments of the relative locations of the items on the issue continuum and their relationship to his felt position on the issue. Thus, the scale of judgment that he focuses on the issue affects his inclination to accept or reject particular items. Consequently, an effective change in expressed opinions should result from judgmental changes that affect the perceived congruence of an item with his general point of view on the issue. If a person with a moderately pro position accepts items extending from neutral to strongly pro, for example, a change limited to his perception of the moderately con items, in which they come to appear extremely con, should not affect his expressed position. But a change that makes the moderately con items appear less extreme or even neutral in his perception may incline him to agree with them. The primary purpose of this study was a test of the effects of such a change in judgments on expressed opinions.

METHOD

An after-only arrangement involving three treatment groups was employed. The Wang-Thurstone

¹ The research reported in this article was completed under Grant M-1351 from the National Institute of Mental Health.

scale of attitude toward the treatment of the criminal was modified to refer to delinquents, by substituting the word "delinquent" for "criminal" and making some minor changes in the wording of a few items.² One group, to be referred to as the "no standards" group, judged each of the 20 items of Form A of the modified scale on a 1-11 basis, in terms of its degree of approval or disapproval toward the use of punishment. A second group, to be referred to as the "end standards" group, made the same judgments but were provided with the following comparison standards for the two extreme categories (1 and 11): "Delinquents should only be treated with sympathetic understanding, guidance, and re-education; for the punishment of misguided youngsters is brutal and degrading"; and "All delinquents should be given long-term prison sentences and very severe physical punishment." Under this condition, it was expected that assimilation toward the end standards would occur for nearby items (Hovland, Harvey, & Sherif, 1957; Sherif, Taub, & Hovland, 1958) and perhaps contrast for the others, with the primary net effect being diminished discriminability among items in the middle region of the scale. The third group, to be referred to as the "middle standard" group, was provided with the following statement to be used as the standard for the midpoint of the scale: "Some delinquents should be punished; others should be given another chance." Although it was uncertain what effects an explicit middle anchor would have on judgments of opinion statements, the possibility of a resultant increase in central tendency led to the constitution of such a group. After making their judgments, all Ss responded to the modified Form B in terms of their own opinions on the treatment of delinquents. The two categories of response were agree and disagree. A few general judgmental items completed the questionnaire.

The instructions stipulated that a rating of 1 was to be used for an item that seemed to express a very unfavorable opinion concerning the use of punishment, a rating of 11 for one indicating a very favorable opinion, and a 6 if an item seemed neither predominantly favorable nor unfavorable but appeared to signify neutrality or doubt concerning the use of punishment. The remaining numbers were to be used for statements that fell between either extreme and the neutral point of 6. The Ss were informed that each of the 11 numbers need not be used, for they may feel that none of the 20 statements is a reasonable example of a particular position. They were also cautioned against permitting their own opinions on the topic to influence their judgments of the items. The explicit standards were presented only once, just before the first item in Form A.

² The need for equivalent scales and the knowledge that college students generally adopt a position opposing the use of punishment in the treatment of delinquents led to the selection of this topic.

TABLE 1
MEAN AFTER OPINIONS ON DELINQUENCY

	No Standards	End Standards	Middle Standard
Mean	4.29	4.68	5.22
SD	1.70	2.15	1.99
N	68	59	58

Booklets covering the three treatments were arranged in a random sequence, following a table of random numbers. They were handed out individually to the Ss who were seen during regular class periods at Boston University and American International College.³ The purpose of the study was stated to be an examination of people's judgments of the opinions of others on a controversial topic.

RESULTS AND DISCUSSION

The opinion data were initially analyzed in terms of a 0, 1 system of scoring.⁴ A score of 1 was assigned for a response which, in terms of the sense of the item, signified some favorableness toward the use of punishment, and a 0 to a response opposing punishment or favoring guidance and leniency in the treatment of delinquents. Neutral items could not be scored under this system. Accordingly, the opinion positions of the Ss are based on only 17 items of Form B and can range from 0 to 17. Nine of the items were phrased punitively and eight nonpunitively. The higher the score, the more punitive is the opinion position.

Effect of Standards on Acceptance of Items

The mean after opinion of each group is presented in Table 1. The *F* for groups, based on analysis of variance, has a *p* value of .05.⁵ By *t* tests, only the middle standard group was found to be significantly more punitive than the no standards group ($p < .01$). In view of the method of scoring, this difference might be due to a decrease in the number of nonpunitive items accepted by the middle standard Ss, an increase in the number of punitive items accepted, or both. When the relevant data are examined, no significant differences are found

³ The author wishes to thank Malcolm Klein of Boston University and Richard Sprinthall of American International College for distributing the booklets to their respective classes.

⁴ The change in attitude object as well as the lapse of time between the development of the test and its current use precludes the use of the original scale values other than as an index of order.

⁵ All *p* values are two-tailed.

TABLE 2
PERCENTAGE OF Ss AGREEING WITH EACH OF THE
TWENTY ITEMS OF FORM B

Item	No Standards	End Standards	Middle Standard
6	2.9	6.8	0.0
12	14.7	16.9	20.7
17	73.5	72.9	70.7
2	92.5	96.6	94.7
8	95.6	91.5	86.2
20	82.4	84.7	86.2
5	75.0	69.5	74.1
11	95.6	84.2	87.9
16	94.1	91.5	98.3
14	76.5	74.6	79.3
1	22.1	30.5	36.2
19	92.6	93.2	94.8
4	80.9	76.3	87.9
10	35.3	37.3	44.8
15	10.3	11.9	22.4
7	7.4	6.8	15.5
13	16.2	16.9	27.6
18	4.4	10.2	10.4
3	0.0	1.7	8.6
9	0.0	1.7	6.9

among the groups in acceptance of the eight nonpunitively phrased items. But the mean number of punitively phrased items that the middle standard group accepted (2.36) is significantly higher than that accepted by the end standards group (1.88) or the no standards group (1.74). The *p* values of the between groups *F* and the individual *t* tests were all $< .05$.

These results are supported by an examination of the patterns of group agreement with each of the 20 items of Form B. Table 2 contains the percentage of Ss in each group agreeing with each item. The items have been arranged according to the original Wang-Thurstone scale values, with the first item the most strongly opposed to punishment and the last the most favorable. Although there are no significant differences among the groups on any item, a higher percentage of the middle standard group than of each of the others agrees with the last 12 items. Items 16, 14, and 1 fall in the neutral category, but the remaining nine are on the punitive side of the midpoint. Considering the percentages as repeated measurements and applying an arc sine transformation to them, an analysis of variance was performed involving groups, items, and groups by items interaction as the three sources of variation. The *F* for groups has a *p* value of .01. Individual *t* tests indicate that the middle

standard group has a significantly higher average percentage of agreement in comparison with each of the others ($p < .05$). The first eight items of Table 2 were grouped as non-punitive ones (i.e., their phrasing was such that acceptance indicated opposition to punishment or approval of guidance and leniency) and the last nine as punitive ones. Separate analyses of variance revealed a significant F for groups only for the punitive items. It is the middle standard group which is significantly different from each of the others; t test p values $< .01$.

Thus, the analyses of the opinion data indicate that experience with the given middle standard led to the expression of a more punitive position. This effect was based essentially on differences in acceptance of the punitive items.

Effect of Standards on Variance and Extremity of Judgments

In terms of the theoretical suppositions underlying the research, group differences in expressed opinions should be associated with a particular pattern of judgmental differences. In general, the variance of scale judgments given by the middle standard group should be significantly less than that obtained under the other conditions. But more importantly, the middle standard group should judge the punitive items less extremely than do the other groups.

Based on the standard deviation of each person's scale judgments, the mean standard deviation was computed for each experimental group. Table 3 contains these means. By analysis of variance, the F for groups has a p value $< .05$. By t tests, the mean standard deviation of the end standards group is significantly greater than those of the other groups ($p = .05$ and $< .02$). Although the middle and no standards groups do not differ significantly from each other, the difference between them is in the required direction.

Examination of the distributions of median group judgments presented in Table 4 confirms this finding. As can be seen, the range of scale values is greater for the end standards group (1.9–11.0) than for the no standards group (2.7–9.3), which in turn has a slightly greater range than that for the middle standard group

TABLE 3
MEAN STANDARD DEVIATIONS OF SCALE JUDGMENTS

	No Standards	End Standards	Middle Standard
Mean	3.24	3.54	3.13
SD	0.75	0.76	0.87
N	66	59	58

TABLE 4
MEDIAN SCALE JUDGMENTS OF EACH ITEM OF FORM A

No Standards		End Standards		Middle Standard	
Item	Median	Item	Median	Item	Median
17	2.7	8	1.9	8	3.1
8	2.9	12	2.0	5	3.1
2	3.6	2	2.2	2	4.0
20	4.2	17	2.4	11	4.6
12	4.8	6	2.5	12	6.0
5	5.3	5	3.1	16	6.0
16	5.3	11	4.3	17	6.2
6	5.5	20	4.9	6	6.4
11	5.5	16	6.2	20	6.4
1	6.4	1	6.4	1	6.4
14	6.4	14	6.4	14	6.4
19	6.9	19	7.1	3	6.7
3	6.9	15	8.8	19	6.8
13	7.9	4	9.3	18	6.8
18	8.5	10	9.5	9	6.8
15	8.7	7	9.6	15	6.9
10	8.9	13	9.9	10	7.1
7	8.9	18	10.1	7	7.1
9	9.2	3	10.5	13	8.2
4	9.3	9	11.0	4	9.1

(3.1–9.1).⁶ It is apparent from the data that besides differences in range, the three groups exhibit different distributions of median judgments. Most of the median judgments of the end standards group fall in the three extreme categories at each end of the scale; the midpoint category contains the majority of the median judgments of the middle standard group; and for the no standards group, the moderate categories are the locations of the majority of median scale values. These differences were found to be significant when analyses were made of the frequency data in Table 5. The p value of χ^2 for the entire table is $< .005$; each group is significantly different from each of the other two groups with p values ranging from $< .05$ to $< .005$.⁷

⁶ The Spearman rhos between the rank order based on the original Wang-Thurstone values for attitude toward the criminal and the order derived from the median judgments for each group are: end standards, .96; no standards, .84; and middle standard, .73.

⁷ In terms of median scale judgments, none of the items whose original scale values would place them on

TABLE 5
FREQUENCY DISTRIBUTIONS OF MEDIAN JUDGMENTS
OF FORM A ITEMS

Scale Categories	No Standards	End Standards	Middle Standard
1-3 or 9-11	5	13	3
4-5 or 7-8	11	4	5
6	4	3	12

TABLE 6
MEAN FREQUENCY OF USE OF SCALE VALUES IN
JUDGING FORM A ITEMS

Scale Values	No Standards	End Standards	Middle Standard
1-3	6.38	6.95	6.22
4-5	1.55	1.14	1.24
6	4.83	4.24	5.95
7-8	2.15	1.88	2.28
9-11	5.08	5.78	4.31

When comparisons among the groups are made in terms of mean scale judgments of the nine punitively phrased and eight nonpunitively phrased items, significant F values for groups are obtained on each type of item (p values $< .05$). The means for the end standards groups were more extreme (i.e., smaller for the nonpunitive items and larger for the punitive ones) than were the comparable ones for the no standards group, whose means were more extreme than were those of the middle standard group.

Evidence for an Assimilation Effect

The data on judgments indicate that standards falling within or near the limits of a range of stimuli exert an assimilation effect. That is, judgments of items near to the standards shift in the direction of the standards. Although the actual scale values of the standards relative to those of the judged items were not obtained, over 80% of the groups provided with the standards agreed, in response to a

one or another side of the midpoint was transposed by any of the groups to the other side of the midpoint. These results are consistent with those obtained by Fehrer (1952). As she noted, when judgments are made of opinion items along a commonly used continuum, gross directional cues can be derived from the customary meanings and usages of language. For example, the directional cue afforded by the phrase "must be punished" in the statement "delinquents must be punished" is so unambiguous that transposition to the lenient side of the midpoint is not likely to occur.

multiple choice question, that they validly represented the designated scale categories.

To obtain some information on the empirical basis for the assimilation effect, the frequency with which each scale category was used was examined. For it may be supposed that assimilation, when it occurs, is due simply to an increase in the use of the particular category values given to the standards.

When analyses were made of the frequencies with which each scale value was used, only the neutral value of 6 yielded a significant F for groups ($p < .05$). The middle standard group used this category significantly more times than did each of the other two groups. Although differences in the use of the other categories were not reliable, there were certain obvious trends to the direction of the differences. To exhibit these, the data were organized in the form of Table 6. Besides the significant F for Category 6, the F for the combined scale values 9-11 had a p value $< .05$; by t tests, the middle standard group used these values significantly less frequently than did the end standards group. The end standards and no standards groups did not differ significantly from each other in the frequency with which they used any single category or combination of categories presented in Table 6. There is, however, an indication that the two sets of extreme categories (1-3 and 9-11) are employed more frequently by the end standards than the no standards groups. When the mean frequencies of use of these six categories are examined, the F for groups is significant ($p < .05$) and the mean of the end standards group is significantly greater than that of the no standards group ($p = .05$).⁸

Although numerous comparisons have been made, the data seem to reveal patterns of use of the scale values that are comparable to the distributions of median values for the 20 items of Form A (see Table 4 and 5). On the basis of these data, the assimilation effect does not seem to result *solely* from an increase in

⁸ The mean number of categories used by each group in judging all 20 items was: end standards, 6.25; no standards, 6.48; and middle standard, 5.64. By analysis of variance, the F for groups was not significant. Thus, the experimental variations affected the distributions of judgments and not the number of categories used in making the judgments.

the frequency with which the scale values of the designated standards are used. Such an increase held under the middle standard condition. But under the end standards condition, although some increase occurred in the use of scale values 1 and 11, it was only when the six end categories were combined that a significant effect was exhibited.

The Relationship Between Judgments and Opinions

In general, the data on group differences in expressed positions and in patterns of scale judgments provide empirical support for the theoretical assumptions underlying the research. A marked reduction in discrimination among opinion items around the more punitive limit of the range of acceptance apparently led to an extension of the range in that direction. Since an overwhelming majority of the Ss in all groups agreed with at least six of the eight nonpunitive items, judgmental effects on these would not be expected to influence agreement with them. The assertion of support for the research hypothesis is based also on the assumption that the effects of the standards perseverated and influenced in a comparable manner judgments of the Form B statements. Considering the relative equivalency of the forms and the very brief delay between responses to them, this assumption seems tenable.

Before data are presented concerning the within-groups relationship between judgments and opinions, some assumptions must be made explicit. It is assumed that judging the 20 items of Form A in the absence of explicit standards had little, if any, effect on opinion responses. It is likely that the Ss' opinions were already crystallized as a result of prior exposure to information and points of view expressed in the mass media, in relevant college courses, and by other people. The topic of juvenile delinquency has been a relatively "live" one over the past few years. It is also assumed that the opinions expressed by the end standards group were not affected by the preceding judgment experience; the difference in mean opinion position between this group and the no standards one did not approach significance. Since the judgmental experience with the middle standard did affect expressed opinions, the relationship between judgment

TABLE 7
RELATIONSHIP BETWEEN Ss' OWN OPINIONS AND DISTRIBUTIONS OF MEDIAN JUDGMENTS OF FORM A ITEMS

Scale Categories	No Standards		End Standards	
	≥5 ^a	<5	≥5	<5
1-3 or 9-11	3	12	6	14
4-5 or 7-8	11	5	8	3
6	6	3	6	3
p_z^2	<.02		<.05	

^a Opinion scores based on responses to Form B.

and opinion for this group is not readily interpretable and, therefore, is not included in this analysis.

On the basis of the 0, 1 system of scoring the opinion responses, Ss were divided into two subgroups: those with an opinion score of 5 or greater (moderate opposition to punishment) and those with a score less than five (strong opposition to punishment). (This cutoff point is slightly above the over-all median of 4.72.) The judgmental patterns of these subgroups were then determined. The data in Table 7 are based on the median scale judgments for the 20 items of Form A. As can be seen, the strongly opposed subgroups place significantly more items in the end categories than do the moderately opposed subgroups. (Approximately equal numbers of items were placed by the subgroups in the 1-3 as in the 9-11 categories.) Rank order correlations were computed between the rankings assigned the items of Form A by the opinion subgroups of the same experimental group. For the end standards subgroups, the correlation is .87 and for the no standards subgroups, .53. Thus, the ordinal arrangement of the items appears to have been affected more by the Ss' opinions when no standard was present than when the end standards were given.

If a person's opinion position on an issue is crystallized and salient, it may function to some extent as an implicit standard and thereby affect his judgments of issue-related stimuli. (In view of the publicity given to the general problem of delinquency in recent years, it is likely that the two required conditions of crystallization and saliency obtained for most Ss.) Since the opinion position of one subgroup is close to the lenient end of the scale and that of the other is nearer to the midpoint, the

distributions of judgments for these subgroups should differ primarily in the number of items located around the middle region of the scale. That is, the distributions should approximate what might be expected if explicit anchors were located on the scale at the mean positions of the subgroups. Furthermore, if opinion position does function as an implicit standard, its effects on judgments should be less when explicit standards are given than when they are absent. This expectation accords with the finding of a lower rank order correlation between the opinion subgroups under the no standards condition than under the end standards one. Also, if it is assumed that the "impressiveness" and saliency of an implicit standard is greater for an extreme opinion position than for a moderate one, then the following prediction can be made: the effectiveness of explicit standards in producing judgmental differences will be less when the comparison groups hold to the same extreme position than when they hold to the same moderate position. The current data provide some support for this inference. The rank order correlation between the median judgments of the two strongly opposed subgroups is .95, but only .68 when the moderately opposed subgroups are compared. In general, the data indicating a within-group relationship between opinions and judgments are consistent with the results of some previous research (Cartwright, 1941; Hovland & Sherif, 1952; Sherif & Hovland, 1953).

SUMMARY

The primary purpose of the research was to examine the effects of changes in the judgments of the scale positions of opinion items on the judges' own opinions on the issue. Three groups were formed by random assignment of subjects. One (the "no standards" group) judged on a 1-11 basis the scale positions of a modification of the Form A items of the Wang-Thurstone scale of attitude toward the treatment of the criminal. The modification involved the substitution of the word "delinquent" for "criminal." The Ss then responded to the similarly modified Form B in terms of their own opinions. Another group (the "end

standards" group) responded similarly but were provided with explicit statements to be used as reference standards for the 1 and 11 categories when judging the Form A items. The third group (the "middle standard" group) differed from the others only in that they were given an explicit reference statement for the midpoint category of 6. Some additional judgmental questions completed the questionnaire.

The results indicate that the explicit middle standard enhanced central tendency and the end standards counteracted it. The mean opinion position of the middle standard group was significantly different from that of the no standards group; but the position of the latter group was not different from that of the end standards group. Thus, a certain kind of change in judgments was found to be associated with a change in opinion position. The differences among the opinion positions of the groups were shown to be due primarily to an increase in acceptance of punitive items by the middle standard group. Under the end standards and no standards conditions, opinions on the issue were found to be related to judgments of the scale positions of the Form A items.

REFERENCES

- CARTWRIGHT, D. C. Relation of decision-time to the categories of response. *Amer. J. Psychol.*, 1941, **54**, 174-196.
- FEHRER, ELIZABETH. Shifts in scale values of attitude statements as a function of the composition of the scale. *J. exp. Psychol.*, 1952, **44**, 179-188.
- HOVLAND, C. I., HARVEY, O. F., & SHERIF, M. Assimilation and contrast effects in reactions to communication and attitude change. *J. abnorm. soc. Psychol.*, 1957, **55**, 244-252.
- HOVLAND, C. I., & SHERIF, M. Judgmental phenomena and scales of attitude measurement: Item displacement in Thurstone scales. *J. abnorm. soc. Psychol.*, 1952, **47**, 822-832.
- SHERIF, M., & HOVLAND, C. I. Judgmental phenomena and scales of attitude measurement: Placement of items with individual choice of number of categories. *J. abnorm. soc. Psychol.*, 1953, **48**, 135-141.
- SHERIF, M., TAUB, D., & HOVLAND, C. I. Assimilation and contrast effects of anchoring stimuli on judgments. *J. exp. Psychol.*, 1958, **55**, 150-155.

Received January 16, 1958.

STRUCTURAL ANALYSIS OF DREAMS:

CLUES TO PERCEPTUAL STYLE¹

BERNICE T. EIDUSON

Hacker Foundation for Psychiatric Research and Education

RECENT experimental work (Klein, 1951) has supported the hypothesis that general cognitive styles or organizing principles of perceiving characterize each individual. These cognitive system-principles seem to operate as personality constants and, in the ways they select, organize, and distort perception, reveal character trends that have emerged in the person's development. Many experiments in this tradition (cf. Adams, 1957; Miller, 1951) have studied perception under subliminal conditions, or in situations where the drive-state has been very intensified, or where fatigue or drugs have weakened integrated functioning. It has been assumed in these experiments that perception occurs simultaneously on many levels of awareness. The question arises whether the same cognitive styles that have been described mainly under normal waking cognition are found on different levels of consciousness. It is to this question that the present investigation was directed.

This study proposed to test whether persons who show differences in the perceptual attitude of flexibility-rigidity as estimated from the Rorschach test also differ predictably along this dimension in the perceptual response pattern of their dreams. The dimension of flexibility refers to the ability to vary, shift, and alter the response pattern. It implies an openness to change, a tolerance for free fantasy, and for lability in emotion, a tendency toward diversity in ways of expression. Its polar opposite is rigidity, which reflects a tight defense against variation or change, a reluctance in experimenting with the new, and a complacency with the original perceptual response. Empirical studies by Klein (1951) and others (Klein & Schlesinger, 1951) have yielded increasingly impressive indications that this dimension comprises one of the regulatory principles in cognition. Our own preliminary work has suggested that the flexibility-rigidity

dimension was appropriate to performances both on the Rorschach test and in dreams.

The Rorschach and dream situations were selected for perceptual study for two reasons: first, the states of psychological awareness that they represent are different—the dream reflects functioning when defensive and integrating mechanisms are relaxed, with the usual spatiotemporal, reality-oriented cues abandoned, except as they are instituted in reporting the dreams; the Rorschach occurs in the waking state and expresses the highly individualistic primary modes of perception only as they are subordinated to the logical, rational framework that the test situation demands. Secondly, both are perceptual situations that occur naturally in the clinical therapeutic experience. We were interested in seeing whether clinical material could be felicitously used in perceptual studies and perhaps provide a source of data not normally available in the laboratory situation.

METHOD

Rorschach Criteria

Rorschach records of 18 outpatients who were being seen in intensive treatment at a psychoanalytically oriented clinic were reviewed by three clinical psychologists working independently. The Rorschachs had been administered as part of the diagnostic study prior to treatment and independent of this study. The psychologists were instructed to consider the test material globally and to rank each subject in terms of the hypothesized perceptual variable, using the following *a priori* criteria: number of major determinants used simultaneously, amount of spread in use of possible determinants for each response, variety of major determinants used in the first response, variability in content categories, variability in sequence, degree of elaboration of responses, degree of constriction of record, number of responses to each card. Judges were free to decide on the relative importance of these criteria in each case. The coefficient of concordance, corrected for continuity, was applied to provide the estimate of agreement among judges. The *F* ratio obtained was 9.5; the *W'* was .76, which indicated that the three sets of ranks showed agreement beyond what might be produced by sampling variance. A composite ranking for the group was made based on the sum of ranks. From the ranking of the total group, the six sub-

¹ Presented in part at the XVth International Congress of Psychology, Brussels, 1957.

jects at each end of the ranking were selected to form the population for this investigation.

Subjects

The total group was composed of four men and eight women ranging in age from 23 to 37. Two men and four women were rated as "flexibles," and their range in age and educational background was roughly similar to the six "rigids." All had been diagnosed as presenting neurotic or character disorders. All of the Ss had been in treatment from 6 to 18 months prior to the start of this investigation. They were known to be persons who were so intensively involved in treatment that they were producing dreams regularly during their therapeutic hours. The patients were unaware that they were Ss, and there has been no indication that the process of their psychotherapy was affected in any untoward way by their being used as research Ss without their knowledge.

Dream Protocols

Psychotherapists were instructed to record verbatim all dreams produced during course of treatment in the five-week period of this study. Since the psychotherapists had been accustomed to collecting dreams as part of their regular therapeutic procedure, little variation in technique was required. Some therapists recorded in their notebooks as customarily done; others used forms provided by the investigator. All dream material occurring during one night was considered as a single dream unless the S spontaneously mentioned awakening between dreams. A mean of 13 dreams per S (145 dreams in all) were collected.

New techniques of analysis had to be adopted for treatment of the dream material for the present purpose. Classically dream analysis (Freud, 1950; Hall, 1951; Leavitt, 1957) has dealt only with the content of imagery and its expression through the mechanisms of displacement and condensation. The concrete image was taken as the scorable response in order to avoid any interpretation of content. Previous study of sensory modalities employed in dreams (Freud, 1950; Knapp, 1956) had pointed preponderantly to their visual character. In our work, however, we were interested in the readiness for fluctuation, shift, and variability as it was revealed in the sense imagery.

The concrete dream imagery was rated on the following categories that were presented on check list forms to psychotherapists:

- a. Sense modalities employed—(visual, auditory, tactual, kinesthetic, olfactory)
- b. Degree of intensity of imagery—(rated on a five-point scale from "very intense or vivid" to "faint")
- c. Amount of affect accompanying imagery—(three-point scale: none or minimal; some or moderate; great or strong)
- d. Kind of affect accompanying imagery—(five-point scale varying from "very pleasant" to "very unpleasant")

Psychotherapists were given criteria for rating of dream imagery and were given pretraining. At no time, however, were the psychotherapists informed about the

perceptual dimension under investigation. The verbatim recording of dreams permitted independent rating of raw data by the investigator, and all discrepancies of more than one scale point were discussed and re-evaluated with the therapist. On a group of 20 dreams, the interrater agreement (psychotherapist and investigator) was .86.

Each dream was rated according to the categories listed above, and then the total set of dreams for each S tabulated. Totals were converted to percentages to allow for difference in the number of dreams reported by Ss during the period of investigation. Then, to provide an estimate of flexibility or rigidity in dream imagery, the percentages were analyzed in terms of the following *a priori* criteria:

- A. Total number of sense modes used simultaneously. (Percentage of times 1, 2, 3 modes used together in the S's series of dreams)
- B. Incidence of variance in primary mode in which imagery was expressed. (Percentage of times primary mode was visual; percentage of times primary mode, auditory; etc.)
- C. Spread of modes when visual mode was primary one used. (Percentage of times visual was accompanied by auditory; by tactual; by auditory and tactual; etc.)
- D. Variation in intensity with which images were expressed. (Percentage of times there was a change in degree of intensity from that which S used most frequently, the modal response)
- E. Variation in amount of affect attached to imagery. (Change from average [mode])
- F. Variation in kind of affect attached to imagery. (Change from average [mode])

RESULTS

The experimental hypothesis was tested by correlating each S's rank on the flexibility-rigidity dimension in the dream imagery with his rank on the dimension as evaluated from his Rorschach test performance. In order to obtain the Ss' ranks on dream imagery, first the nonparametric sum of ranks was applied to each of the above criteria to test the significance of the difference between means of Ss who had been grouped as "rigid" or "flexible" on the basis of their Rorschachs. The results showed all but one of the variables to be significant (Table 1). Therefore, Ss who had been rated as "flexible" on their Rorschachs were generally rated as "flexible" when the categories for flexibility in dream imagery were considered individually. Then with the median scores of the total group of 12 Ss as the cut-off points in each of the categories, each S's performance was ranked in the five categories which had shown significance. The final rank ordering of Ss on the flexibility-rigidity dimension

sion in dream imagery was obtained by taking the average (mean) of these ranks.

A rank correlation coefficient of .74, significant at the .01 level, was obtained when the ratings of flexibility-rigidity in Rorschach test performances were compared with those found in dream imagery. Persons who are flexible in their mode of organizing Rorschach responses thus tend to be similarly flexible in their use of imagery in dreams.

Behavioral Correlates of Flexibility-Rigidity in Dreams. Confirmatory validity for the rating on the flexibility-rigidity dimension in dreams was sought by evaluating whether the "flexible" Ss actually functioned in a flexible way in various aspects of their behavior and whether the "rigid" Ss displayed rigidity in their functioning. Psychotherapists who knew their behavior patterns intimately were asked to rate them in these areas: work, interpersonal relationships, attitudes toward self, imaginative and ideational resourcefulness, and emotional behavior. A four-point scale in which the degrees of flexibility were distinguished was used. Psychotherapists were instructed that flexibility referred to the extent to which the S showed variation, instabilities, shifts, or unsteadiness in his behavior pattern. They were also told that we were interested primarily in the amount of shift or change in behavior rather than in what this variability meant in the patient's personality picture and therefore that a high degree of flexibility could for some patients signify the potential for appropriate and varied responses to different situations, while for others it might mean the lack of directiveness or the unpredictability in behavioral response. This part of the study was done four weeks after the dream collection had been discontinued and was the first time psychotherapists were informed about the perceptual dimension under study.

To obtain a composite rating of "flexible" or "rigid" in behavior for each S, the clinical ratings made in each of the five specific areas were averaged, each having equal weight (Table 2). The mean of the four scalar points (2.5) served as the cutting-off point for the general rating as "flexible" or "rigid" (cf. Table 2.). Using Fisher's exact test on 2×2 contingency tables, the independence of ratings

TABLE 1
RANK ORDER RATINGS OF "FLEXIBLES" AND "RIGIDS"
ON DREAM IMAGERY VARIABLES

Subject Number	Variables ^a					
	A*	B**	C**	D	E*	F*
"Rigid" subjects						
2	6	10	9	7	7.5	2
3	12	8	9	2	12	8
6	9	12	12	12	7.5	11.5
7	6	7	11	11	9	11.5
8	10	11	9	5	5	10
11	11	9	7	9	11	9
"Flexible" subjects						
1	6	4.5	5	1	4	4
4	1	1	1	3.5	3	1
5	2	6	6	10	10	3
9	8	4.5	3	8	2	5.5
10	4	2	4	3.5	1	7
12	3	3	2	6	6	5.5

^a Variable A—Total number of sense modes used simultaneously; B—Incidence of variance in primary mode in which imagery was expressed; C—Spread of modes when visual mode was primary one used; D—Variation in intensity with which images were expressed; E—Variation in amount of affect attached to imagery; F—Variation in kind of affect attached to imagery.

*.05 critical point for smaller rank totals for $n_1 = 6$, $n_2 = 6 = 26$ (White, 1952; Wilcoxon, 1945), computed on only one side.

** .01 critical point for smaller rank totals for $n_1 = 6$, $n_2 = 6 = 23$ (White, 1952; Wilcoxon, 1945), computed on only one side.

of actual function and of dream imagery was tested (Finney, 1948).

The correlation between dream imagery and overt behavioral functioning was significant at the .05 level. This would confirm the hypothesis that flexibility operates with a high degree of consistency in the S's style of perceiving and in his everyday functioning. It suggests that such organizing modes may act as "sets" for behavior. This relationship between dreams and behavioral functioning also assures us that an appropriate measure was developed for reflecting the flexibility-rigidity dimension in dreams.

With respect to psychotherapists' ratings for specific areas, significant lack of independence was found between ratings of flexibility in dream imagery and in attitude toward self. This was the only area that reached the .05 level of significance. However, ratings in areas of interpersonal relationships, emotional behavior, and ideational resourcefulness were significant at the .10 level (when chi square

TABLE 2
RATINGS OF FLEXIBILITY-RIGIDITY IN BEHAVIOR

Subjects	Variables					
	Work	Interpersonal Relationships	Attitudes Toward Self	Ideational Resourcefulness	Emotional Behavior	General Rating on Flexibility Dimension ^b
1	4	3	4	3	4	f
2	3	1	3	1	1	r
3	1	1	1	1	1	r
4	— ^a	3	4	2	3	f
5	3	3	3	3	4	f
6	2	2	3	3	3	f
7	2	4	2	4	4	f
8	4	2	2	2	2	r
9	4	4	4	3	4	f
10	2	4	3	3	3	f
11	3	2	2	2	2	r
12	1	3	4	3	2	f
	$\chi^2 y = .12$ $p > .50$	$\chi^2 y = 2.83$ $p < .10$	$\chi^2 y = 5.18$ $p < .05$	$\chi^2 y = 2.83$ $p < .10$	$\chi^2 y = 2.83$ $p < .10$	$\chi^2 y = 5.18$ $p < .05$

^a No work history.

^b Obtained by getting mean of ratings listed and using mean of scalar points (2.5) as cut-off point.

with a Yates correction was applied), suggesting a tendency toward consistency in these specific areas of behavior and imagery (cf. Table 2).

DISCUSSION

In this study the same cognitive style principle has been shown in behavior on different levels of psychological awareness. In the dream where ego controls are relaxed and the primary process takes dominance, the flexibility-rigidity style principle was found to operate as a mode of organizing perception, as it does in the cognitive test situation administered under normal waking conditions. In the study of overt behavioral functioning the same cognitive style principle was also observed. Because of the limitations presented by this circumscribed procedure, this can at best be a tentative conclusion demanding more refinement and validation. However, the results point to the need for exploring perception under a variety of differing conditions of awareness.

Speculation as to the significance of these findings directs attention to the possibility that precursors of cognitive style patterns that we see in rational logical secondary processes may exist in primary process thinking. Classically, primary thought processes have been described as being mainly unorganized, but the question is whether or not the pre-

cipitates of the organization which ultimately the ego displays are contained therein. The early psychoanalytic position stated that the id, out of which the ego was derived, was completely chaotic and without order; however, psychoanalysis today holds that from the very first the original endowment shows primitive ego regulatory mechanisms, that there is no period in the individual's life when some suggestion of these mechanisms is not found, and that even in reflex activities these are already operative. This study suggests that the primitive ego formation is likely to have structural dimensions, and that in early development these may set the stage for the perceptual response pattern which we see in adult purposeful and directed function.

One further bit of clinical data is provocative. Among the material available on each S was the predominating symptom picture. Inspection of this showed that the symptoms of the Ss in the flexible group were alike in that all were primarily directed out onto the environment: four showed impulsive behavior with sexual acting out or difficulties with authority, and two were paranoid character disorders.

The patients in the rigid group by contrast tended consistently to show symptoms which were directed inwardly, or toward themselves: two, psychosomatic or conversion disorders; three, depressions; one, phobias. The implica-

tions of this alignment are not clear but are suggestive of differences similar to those reported between direction of aggression and certain biochemical correlates—for example, in the work of Funkenstein, King, and Drolette (1954) and others (Eiduson, Crumpton, & Brill, 1957) who have related direction of aggression to the epinephrine-norepinephrine ratio. This relationship points to a fruitful avenue for further research, although the mere generalization of the correlates of hypothesized attitudes does not explain the processes involved.

SUMMARY

This experiment was directed to the question of whether the same perceptual style principles that have been described under normal waking conditions are found on different levels of consciousness. The experimental hypothesis was that persons who show differences in the perceptual attitude of flexibility-rigidity as estimated from the Rorschach test also differ predictably along this dimension in the perceptual response pattern of their dreams. Twelve patients in intensive treatment at a psychoanalytically oriented clinic were administered the Rorschach test, which was rated for degree of flexibility on a number of a priori criteria. From each of the subjects, who were producing dreams regularly during their therapeutic hours, a mean of 13 dreams (145 in all) was collected. Concrete dream imagery was analyzed for flexibility or rigidity according to a number of a priori criteria which were considered appropriate to the material. Results confirmed the experimental hypothesis. Ratings as to the flexibility shown by subjects in five areas of overt behavior and in their functioning in general also bore out the rating

of flexibility or rigidity displayed in dream imagery. The findings indicate that in psychological behavior at varying levels of awareness there is consistency in the operation of the cognitive style principle of flexibility-rigidity, and suggest the importance of further exploration of perception under different conditions of consciousness.

REFERENCES

- ADAMS, J. K. Laboratory studies of behavior without awareness. *Psychol. Bull.* 1957, **54**, 383-405.
- EIDUSON, S., CRUMPTON, E., & BRILL, N. Q. Plasma catechol amines as related to direction of aggression in human beings. *Proc. 15th Int. Cong. Psychol., Brussels*, 1957.
- FINNEY, D. J. The Fisher-Yates test of significance in 2×2 contingency tables. *Biometrika*, 1948, **35**, 145.
- FREUD, S. *The interpretation of dreams*. A. A. Brill (Trans.) New York: Random House, 1950.
- FUNKENSTEIN, D. H., KING, S. H., & DROLETTE, M. The direction of anger during a laboratory stress-inducing situation. *Psychosom. Med.*, 1954, **16**, 404.
- HALL, C. S. What people dream about. *Sci. Amer.* 1951, **184**, 60.
- KLEIN, G. S. The personal world through perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality*. New York: Ronald, 1951.
- KLEIN, G. S., & SCHLESINGER, H. J. Perceptual attitudes of "form-boundedness" and "form-lability" in the Rorschach. *J. Pers.* 1951, **19**, 289.
- KNAPP, P. H. Sensory impressions in dreams. *Psychanal. Quart.* 1956, **25**, 325.
- LEAVITT, H. C. Teleological contributions of dreams to the waking ego. *Psychoanal. Rev.*, 1957, **44**, 212.
- MILLER, J. G. Unconscious processes and perception. In R. R. Blake & G. V. Ramsey (Eds.), *Perception: An approach to personality*. New York: Ronald, 1951.
- WHITE, C. The use of ranks in a test of significance for comparing two treatments. *Biometrics*, 1952, **8**, 33.
- WILCOXON, S. Individual comparisons by ranking methods. *Biometrics Bull.* 1945, **1**, 80.

Received January 27, 1958.

AUTONOMIC RESPONSES AND THEIR RELATIONSHIP TO RACE ATTITUDES¹

FRANK R. WESTIE AND MELVIN L. DE FLEUR

Indiana University

THIS is an experimental study of autonomic responses of prejudiced and unprejudiced persons viewing interracial situations. Changes in galvanic skin resistance, finger pulse, duration and amplitude of heart beat, and duration of heart cycle were recorded. The relationship of these measures to race attitudes as indicated both by scales and by open-ended interviews suggests that attitude-response syndromes include emotional tensions and affective feeling tones not under the individual's direct voluntary control or subject to his conscious examination. Experimental exploration of this more subtle dimension of attitude behavior may lead to a clearer understanding of the concept "attitude."

Most direct approaches to the assessment of attitudes rest upon at least two tacit assumptions: (a) that people can introspectively assess what their attitudes are (or "would be" under some hypothetical circumstance) toward the attitude object in question, and (b) assuming that people *do* know what their attitudes are, that they are both willing and able to communicate these faithfully to the investigator. This foundation of assumptions, upon which modern structured attitude scaling rests, stands in need of critical examination. We question whether all attitudinal phenomena exist for subjects on such a verbalized level of awareness as to permit them to look into themselves, and then say, in effect, "this is how unfavorable I am." If this aspect of a person's image of himself is unclear, there is scant possibility that it can be effectively communicated to the investigator, regardless of his ingenuity in constructing orderly scales. Furthermore, if attitudes include deep-seated emotional orientations, respondents may be incapable

altogether of either perceiving or communicating such information. Measurement of such emotional orientations is not possible with traditional paper-and-pencil techniques.

This is not to say that structured scales and paper-and-pencil techniques are useless. On the contrary, we have used them extensively in the present research. The point is rather that they do not yield the whole picture of attitudinal behavior, but provide an image of only one dimension, the verbal.

We find it useful to conceive of attitudinal phenomena as a multidimensional complex, without any assumption of a one-to-one correspondence between the dimensions. One dimension is the emotional orientation we seek to examine; another is the verbalized behavior; still a third dimension includes overt approach-avoidance acts which to many seem to be the "real" behavior that attitude scales should predict. All of these are actual behaviors and constitute important aspects of the attitude complex. Such phenomena exist as recurrent behaviors, and the task of attitude research, as conceived here, is a search for consistency of response within these dimensions.²

Although existence of such dimly perceived feeling states is not difficult to demonstrate, their measurement proves to be a somewhat more exacting task. For example, many persons who have had experience in interviewing white people about their attitudes toward Negroes have felt a shortcoming in standard paper-and-pencil techniques. Individuals are

¹ We wish to acknowledge our indebtedness to the Social Science Research Council for financial assistance and to Rolland C. Davis of the Indiana University Psychology Department who gave generously of both his time and equipment in connection with the physiological measurements. The recording of physiological responses was partially supported by funds under contract No nr 908 (03) between the Office of Naval Research and the Indiana University Foundation.

² The reviews of the literature of the past thirty years by Allport (1935), Campbell (1950), and Green (1954) all stress that one characteristic is almost uniformly mentioned in attitude definitions: consistency of behavior, either in action toward the attitude object or in verbalizing about it. This usage, in effect, equates the term attitude to a probability (of recurrent behavior). Campbell's definition, as paraphrased by Green, expresses this focus: "An individual's social attitude is an enduring syndrome of response consistency with regard to a set of social objects. Many theorists more specifically include in their definitions some statement concerning 'emotional overtones,' 'affective feeling states,' or some similar phrase."

frequently encountered who are determined to be "big" about the "Negro question," not only to the interviewer but to themselves. The strictly verbal picture that emerges from such interviews is that of a person who is generally favorable in his attitudes toward Negroes. But more subtle responses speak eloquently of another picture—a certain uneasiness, the clenched fist, the restless stirring, the new atmosphere of forced friendliness—all of this following closely upon a pleasant, relaxed easygoing atmosphere built around a discussion of an affectively neutral topic. We have observed that this change occurs when the first question about Negroes is asked. Although the interviewer may be quite aware of these subtle cues, he ordinarily lacks techniques with which to control, quantify, or record such responses.

Considerations of this sort led the writers to embark upon a study of emotional tensions in situations of race relations, as indicated by autonomic responses. The relationship of such responses to measures of race attitudes obtained by more structured methods may allow meaningful exploration of this neglected dimension of attitudes.

The problem may be stated as follows:

1. Are there consistencies in autonomic response associated with exposure to objects toward which persons have (scale-measured) attitudes?
2. If there are measurable associations between autonomic responses and attitudinal involvement, what is the direction of the (scale-measured) attitudinal involvement? Do those persons most unfavorable in their verbal response to the attitude-object manifest more pronounced autonomic responses than those more favorable in their verbal responses? Or is the reverse true?

METHOD

The investigating involved three phases: the attitude testing program, the measurement of autonomic responses to attitude stimuli, and the postexperimental interview.

The Attitude-Testing Program

Subjects (Ss) were selected on the basis of scores on the Summated Differences scales.³ These scales re-

³ For a detailed description of the manner in which these scales were constructed, their reliability coefficients, etc., see Westie (1953).

quire the S to indicate the degree to which he is willing to associate (in a variety of social settings) with Negroes of a variety of socioeconomic levels as compared to whites of comparable socioeconomic levels. The total score, based on the summated differences between responses to whites and Negroes of the same socioeconomic level, provides an index of positive or negative verbal attitudes toward Negroes. This battery of scales, made up of some 500 items, was administered to 250 freshmen students at a large university. From this population, a control group of 23 Ss was matched in over-all frequency distribution with an experimental group of the same size in terms of age, sex, social class, residential history, and previous contact with Negroes. The experimental ("prejudiced") group was selected from students who scored in the upper 25% of the Summated Differences scale, while the control ("unprejudiced") group was drawn from those scoring in the lower 25% of the distribution. Half of each group were males, the other half females.

The Laboratory Study

In the laboratory study, the Ss of each group were exposed individually to a series of colored photographic slides depicting whites and Negroes in various situations. During these presentations the S was connected to apparatus designed to measure GSR (galvanic skin response), FP or finger pulse volume (measured by plethysmograph), amplitude and duration of heartbeat (measured by electrocardiograph), duration of heart cycle (time required for the heart to pass from one state to the same state again, as measured by cardiographometer). The physiological measures required about an hour per S. The data were gathered during a 12-week period.

The colored slides were posed and photographed specifically for this study. They were projected through a port into a small air-conditioned room where the S sat relaxed in a chair. Each slide portrayed two "good-looking" young adults sitting side-by-side looking at each other with a pleasant, friendly expression. One slide was included for each possible combination of males and females, whites and Negroes. In addition to these pair slides, a set of four single-person slides was included, portraying a single member of each race and sex. The age of the person photographed approximated that of the Ss. The models were dressed in dignified apparel; the background behind the seated persons included a small table with a lamp and a partial view of a window with a drapery. The white and Negro models appearing in the slides were not known to the Ss. The slides were thus designed to avoid the arousal of strong autonomic reactions for reasons extraneous to the problem at hand.

A suitable rest and standardization period was allowed for each S after being attached to the apparatus. Each S was then shown the 12 stimulus slides. The four single-person slides were shown as a first set, followed by eight pair slides. For both the first set and the second set, an independently drawn random order of presentation was used for each S so as to control for serial effects. Each slide was shown for 30 seconds with a 30-second rest between each presentation.

Although five separate autonomic physiological

responses to these slides were recorded, the present paper considers only two of these: the GSR, recorded through electrodes placed on the palm, and finger pulse volume (FP), registered by an electrode at the end of the index finger. An index of change in GSR patterns was devised which states changes in skin resistance as a percentage of the resistance at the beginning of each stimulus presentation. Since the GSR activity was permanently recorded by a stylus on a recording drum, it was possible to synchronize such changes with the exact time of stimulus presentation. A similar system was used for recording FP changes. Amplified FP activity was permanently recorded on a separate drum paper, and synchronization of this record with the presentation of the stimuli was also possible.⁴

The Post-experimental Interview

Each *S* was interviewed immediately following the measurement of his physiological responses. During this hour-long session, which took place in an office far removed from the laboratory, the *S* responded to a variety of structured and open-ended questions intended to help in giving meaning to the physiological responses. Projective interpretations were elicited to slides comparable to those he viewed in the laboratory, and topics such as the *S*'s previous experiences with Negroes, his personal and social background, his reference groups, and his degree of willingness to be photographed with Negroes were explored.

RESULTS

Quantified variations in these physiological responses were arranged in a factorial analysis of variance design. The statistical analysis was complicated by the fact that each set of slides (single person slides vs. paired slides), and each physiological measure yielded separate data.

In addition, for the GSR, the continuous kymograph records provided several ways of obtaining measures of the results. Thus, GSR responses to the single-person slides were analyzed on the basis of measures made during the *first third* of the 30-second period during

which each *S* viewed each slide. Similarly, the *last two thirds* of each stimulus presentation provided the basis of another complete analysis. Still another analysis was made of changes in autonomic activity during each 30-second stimulus period taken as a unit. From these various combinations, six analyses were completed.

In the present paper only the analysis of the last two thirds of the stimulus period is presented. Although the results of the different analyses are quite similar, our data seem to indicate that fullest response occurred during this last 20-second period of the stimulus presentation.⁵

Table 1 shows the general statistical design and gives the mean level of finger pulse (FP) response to the single-person slides by various categories of *S*s to the different types of slides. In addition, the usual summary table for the analysis of variance, based on the FP responses to the single-person slides, appears in Table 2. A verbal summary of the significant sources of variance in all unreported analyses is given in Table 5.

Examination of Table 1 indicates that FP activity was generally less for females than for males, but Table 2 reveals that the difference is not significant. Similarly, the data in Table 1 show that prejudiced *S*s had generally lower levels of FP response to the Negro slides than the unprejudiced *S*s, but, again, the difference is not significant. The reaction patterns of these two *S* groups to the white slides were very similar. However, statistically significant results are evident when both *S* sex and *S* attitude are taken into account jointly. This finding can be more clearly understood by com-

⁴ Space does not permit discussion of details concerning the apparatus, the recording devices, the nature of the laboratory, methods of translating response records into quantitative indices, etc. The apparatus is described in detail in Davis, Siddons, and Stout (1954). The procedure employed in the quantification of the kymograph records was relatively simple. For the finger pulse records, the greatest amplitude of finger pulse volume for a 5-sec. period (2.5 to 7.5 sec. after onset of stimulus) was divided by the greatest amplitude of FP volume for a 5-sec. interval preceding the onset of stimulus.

The GSR measures were obtained by taking the maximum peak during a response period as a proportion of the response amplitude before the onset of the stimulus. Simple millimeter measures of the records were used in both cases.

⁵ The remaining analyses, including four variance summaries and four tables of means, have been deposited with the American Documentation Institute. Order Document No. 5848, from ADI Auxiliary Publications Project, Photoduplication Service Library of Congress, Washington 25, D. C., remitting \$1.25 for 35 mm. microfilm, or \$1.25 for 6 by 8 in. photocopies. Make checks payable to Chief, Photoduplication Service, Library of Congress.

The analysis followed that suggested by Edwards (1950) for the situation where repeated observations are made on several groups of *S*s. The interaction between individual *S*s and the stimulus conditions within a given group was taken as the error term for comparison between sets of columns. Mean squares based on row comparisons are divided by a combined "between subjects" term to obtain the appropriate *F* value.

TABLE 1
MEAN LEVELS OF FINGER PULSE (FP) RESPONSE TO THE SINGLE-PERSON SLIDES

Subject Variables	Stimulus Variables					
	White Slides			Negro Slides		
	White Female	White Male	Mean for White Slides	Negro Female	Negro Male	Mean for Negro Slides
Male: Prejudiced	.755	.953	.854	.953	.978	.966
Unprejudiced	.924	1.690	1.307	1.135	.970	1.052
Female: Prejudiced	1.224	1.376	1.300	1.744	1.252	1.498
Unprejudiced	.990	1.018	1.004	.873	.928	.900
Mean for males	.840	1.322	1.081	1.044	.974	1.009
Mean for females	1.107	1.197	1.152	1.309	1.090	1.200
Mean for prejudiced	.989	1.164	1.077	1.348	1.115	1.232
Mean for unprejudiced	.957	1.354	1.115	1.004	.949	.976

Note.—Lower numerical values indicate greater physiological response (see footnote 6). Relative response ranges can be interpreted as follows: Response levels before each stimulus presentation equal 1.000. A value such as 1.744 (prejudiced females responding to Negro female slides) indicates a 74% response change in the direction of decreased FP activity.

TABLE 2
ANALYSIS OF VARIANCE FOR FINGER PULSE RECORDS, SINGLE-PERSON SLIDES

Source of Variation	df	Mean Square	F
Between subjects	33		
a. Between subject types:	3		
1. Between sexes	1	.83	.929
2. Between attitude types	1	.46	.515
3. Sex \times attitude interaction	1	4.26	4.770*
b. Between subjects of same type	30		
1. Prejudiced males	7		
2. Unprejudiced males	7		
3. Prejudiced females	9		
4. Unprejudiced females	7		
Within subjects	102		
a. Between stimulus conditions	3		
1. White vs. Negro slides	1	.00	.000
2. Male vs. female slides	1	.11	.222
3. Race \times sex interaction	1	1.67	3.370
b. Interactions, subjects, and stimulus conditions	99		
1. Interaction between subject, types and stimulus conditions	9	.36	
2. Pooled interaction between individual subjects and stimulus conditions	90	.495	
Total	136		

* Indicates significance at or beyond .05 level.

both sexes, the greatest degree of FP activity was shown by prejudiced males while viewing white female slides, and the smallest response was by prejudiced female Ss viewing Negro female slides.

The analysis of the GSR means (Table 3) for the single-person slides reveals that the prejudiced group tended to give greater levels

paring the "mean for white slides" with the "mean for Negro Slides" for each of the four S types (Table 1). Prejudiced males and prejudiced females both gave smaller FP reactions to the Negro slides than to the white slides.⁶ On the other hand, unprejudiced males and unprejudiced females showed no such decrease in level of FP response toward Negro slides, actually showing a higher level of FP activity. This comparison suggests an inverse relationship between degree of prejudice and autonomic response as measured by FP activity. This relationship is complicated, however, by the fact that the two sexes reacted generally with somewhat different levels of activity. That is, while the direction of response was similar for prejudiced persons of

⁶ Unusually high numerical values found in such cases as prejudiced females viewing Negro female slides (Table 1), need careful interpretation. The 1.744 numerical value represents a 74% decrease in FP activity over the preceding 30-sec. rest period. This seems difficult to account for in terms of relaxation of the arterioles of the finger. Rather, it would seem that the stroke volume of the heart had increased to a point where it more than cancelled the constriction response. If this is the case, the suggested inverse relationship between verbal attitude and FP activity may prove with further research to be a more complicated product of heart action and arterial constriction; a product whose components are each related to the verbal attitude variable in a different way. Independent measures of heart stroke were attempted in the present research but proved unusable. Further research on this problem should give careful consideration of this possible alternative interpretation of the suggested inverse relationship between FP activity and attitudinal predisposition.

TABLE 3
MEAN LEVELS OF GSR RESPONSE, SINGLE-PERSON SLIDES, LAST $\frac{2}{3}$ OF STIMULUS

Subject Categories	Stimulus Conditions					
	White Slides			Negro Slides		
	White Female	White Male	Mean for White Slides	Negro Female	Negro Male	Mean for Negro Slides
Males: Prejudiced	.948	.981	.965	.932	.957	.940
Unprejudiced	.920	.934	.927	.840	.901	.870
Females: Prejudiced	.901	1.058	.980	.886	.953	.920
Unprejudiced	.928	1.111	1.020	.892	.879	.886
Mean for males	.934	.958	.946	.886	.929	.905
Mean for females	.915	1.085	1.000	.889	.916	.903
Mean for prejudiced	.925	1.020	.973	.909	.955	.930
Mean for unprejudiced	.924	1.022	.973	.866	.890	.878

of autonomic response to the Negro slides than did the unprejudiced group, suggesting a *direct* relationship between prejudice and GSR. In each of the three parallel analyses previously mentioned, a similar pattern of significant results appeared.

Table 3 also reveals that GSR responses by all *S* types were consistently greater to slides portraying males than to those showing females. The same situation prevails in all three analyses of the single-person slides.

In summary of the FP and the GSR analyses of the single-person slides, prejudiced *Ss* manifested different levels of autonomic activity than unprejudiced *Ss* while viewing Negro slides as compared to white slides. Prejudice and FP activity tended to be inversely related, while prejudice and GSR activity tended to be directly related. In addition, the sex of the *S* influenced the level and direction of response.

We turn now to a summary of the analyses of the pair slides, restricting our discussion to the GSR data.

Table 4 shows the mean levels of response toward the various types of slides. A significant *F* value (see Table 5) was obtained for GSR reactions to slides portraying persons of the "same sex" versus those presenting "mixed sex" pairs. The means of the lower part of Table 4 show that *Ss* gave greater responses to slides portraying persons of the same sex.

It should be kept in mind, however, that by the time the pair slides were presented to the *Ss*, they had already reacted to the set of single-person slides. The *Ss* tended generally

to respond with lesser magnitude, indicating a "jading" effect had set in by the time that the pair slides were presented.

The statistically significant findings of the experiment can be listed briefly as follows:

Finger Pulse Analysis

1. The degree of autonomic response accompanying exposure to the slides was related to the sex and attitudes of the *Ss* taken in combination. The response of greatest magnitude was by prejudiced males viewing white female slides. The response of smallest magnitude was by prejudiced females viewing Negro female slides (Table 1 and 2).

2. Prejudiced *Ss* of both sexes showed lesser levels of FP activity while viewing Negro slides as compared to white slides, while unprejudiced *Ss* of both sexes showed greater reactions to Negro slides as compared to white slides, suggesting an inverse relationship between FP activity and prejudice (Tables 1 and 2).

Galvanic Skin Response Analysis

1. Prejudiced *Ss*, in the single-person slide analysis, showed greater responses toward Negro slides than did unprejudiced *Ss* (Table 3).

2. The *Ss* as a whole, regardless of sex or attitude, gave consistently greater responses to slides portraying males than to slides portraying females (Table 3).

3. In the analysis of the pair slides, the *Ss* as a whole showed higher levels of autonomic

TABLE 4
MEAN LEVELS OF GSR RESPONSE, PAIR SLIDES LAST $\frac{2}{3}$ OF STIMULUS

Subject Variables	Stimulus Variables					
	Same-Race Slides			Mixed-Race Slides		
	Same Sex	Mixed Sex	Mean for Same Race	Same Sex	Mixed Sex	Mean for Same Sex
Males: Prejudiced	1.030	1.037	1.033	1.256	.935	1.096
Unprejudiced	1.046	.979	1.013	1.015	.949	.982
Females: Prejudiced	1.004	1.017	1.001	.962	.993	.976
Unprejudiced	.953	.946	.950	1.173	.932	1.053
Mean for males	1.038	1.008	1.023	1.136	.942	1.039
Mean for females	.979	.982	.981	1.068	.963	1.015
Mean for prejudiced	1.016	1.027	1.022	1.109	.964	1.036
Mean for unprejudiced	1.000	.963	.982	1.094	.941	1.018
Same-sex slides			Mixed-sex slides			
Mean for prejudiced			1.062			.991
Mean for unprejudiced			1.042			.952
Mean for males			1.037			.925
Mean for females			1.023			.972

TABLE 5
SUMMARY OF SOURCES OF VARIANCE FOR ALL ANALYSES

Physiological Variable	Slide Stimuli	Stimulus Period	Significant* Source of Variance	Direction of Relationship
FP	Single-person slides	First $\frac{1}{3}$ of period		(See Tables 1 and 2)
GSR	Single-person slides	First $\frac{1}{3}$ of period	White vs. Negro slides	Average S's response greater to white slides with prejudiced S's response greater to Negro slides
			Male vs. Female slides	Response greater to male slides for all S types
GSR	Single-person slides	Last $\frac{2}{3}$ of period	White vs. Negro slides	Average S's response greater to white slides with prejudiced S's response greater to Negro slides
			Male vs. Female slides	Response greater to male slides for all S types
GSR	Single-person slides	Entire period	White vs. Negro slides	Average S's response greater to white slides with prejudiced S's response greater to Negro slides
			Male vs. Female slides	Response greater to male slides for all S types
GSR	Pair slides	First $\frac{1}{3}$ of period	none	
GSR	Pair slides	Last $\frac{2}{3}$ of period	Same-sex vs. mixed-sex slides	Response greater to same-sex slides for all S types
GSR	Pair slides	Entire period	none	

* Significant at the .05 level.

response toward slides portraying persons of the same sex as compared to mixed sex pairs (Table 4).

DISCUSSION

How do these findings bear on the problem as stated originally? The most simple and general answer that can be given to this question is that responses to attitude objects include nonrandom autonomic physiological patterns that are related to characteristics of the *S*, characteristics of the object, and combinations of these characteristics. There is evidence that the attitudinal predispositions of our *Ss*, as measured by our attitude scales, were related to some extent to their autonomic response patterns. There is also ample evidence that the response patterns were complicated by the influence of other variables. The sex category of the *S* operated in various ways to determine partially the direction and magnitude of response. In addition, a person's autonomic responses were influenced by the racial and sexual categories of the attitude object, and the ways in which attitude objects were paired together.

What do these data mean in terms of traditional categories of affective orientation? It cannot be assumed that conspicuous physiological changes manifested by an *S* upon being exposed to a stimulus object indicates simple favorability or unfavorability toward the attitude object, regardless of his scale-measured attitudes toward that object. It is easy to fall into thinking, especially when the stimuli are "race objects," that high response indicates a negative attitude. Actually, such responses could be due to a number of conditions. Conflicting feelings, indecision, or even positive involvement might result in large autonomic responses. There are probably many other interpretative possibilities besides these. The present research makes no attempt to assign simple meaning to the responses. Our aim was to determine if measurable patterns of autonomic physiological response exist and, if so, are they related to scale measured attitudes?

Finally, in attempting to evaluate the experimental data, two aspects of our method should be kept in mind. First, it would have been easy to design the slide stimuli so that larger autonomic responses were evoked. It

will be recalled that our slide stimuli were deliberately made "conservative." That is, suggestive poses, body contact, unusual clothing, or facial expressions were deliberately avoided, so that if autonomic responses occurred it came in spite of the avoidance of these secondary cues. Further explorations of this problem could well include more open portrayal of emotion-arousing interracial scenes. There is every reason to presume that response consistencies under such circumstances would be more clear-cut. Another variable that might well be introduced is the influence of well-defined social roles. For example, the Negro man and white girl portrayed together might evoke entirely different response patterns depending upon whether they were shown lying on a beach together clad in bathing suits, or talking together in a church with the Negro in a clergyman's robes. Such social roles were not included in the present research, but they may importantly influence attitudinal responses.

The second aspect of method which may have placed limitations on the magnitude of response is the order in which the slides were presented. Every *S* viewed the four single-person slides before being presented the pair slide stimuli. It is clear that there was a tendency for GSR activity to become dampened with rapidly presented stimuli. Since the data for the pair slides were collected during the last half of the laboratory session, differences between reactions might tend to be obscured. Further experimentation should use a reduced number of stimuli. One or two preliminaries at the start, followed by not more than four or five should be a maximum.

In terms of their implications for the attitude concept, the over-all results indicate that attitudinal responses include autonomic physiological activity that is related to the attitudinal characteristics of the individual as measured by more standard means.⁷ They are also related to the sex of the responding individual and to features of the object of response.

⁷ A recent article by Rankin and Campbell (1955) analyzes GSR responses of male students to a Negro by the ingenious device of having a Negro experimenter touch the *Ss* while "adjusting" the apparatus. This problem of attributing the response to the racial aspects of the situation vs. a reaction to a particular person is discussed.

Clearly, there is little point in attempting to interpret these physiological responses as the "true attitude." Rather, it seems more useful to regard these responses as another dimension of attitudinal behavior to be considered along with the verbal and overt action dimensions.

Although one need not assume that this dimension must correspond with other dimensions in a one-to-one fashion, it would seem theoretically useful and empirically possible to establish the nature of the relationship between the various dimensions for particular populations under given circumstances. Our findings, however modest, indicate that further research along these lines may prove useful not only for clarifying the construct "attitude," but also for increasing our knowledge of the empirical referents of the construct.

SUMMARY

The problem was to determine if autonomic response was associated with exposure to objects toward which individuals have (scale-measured) attitudes. That is, do persons who are unfavorable in their verbal response to Negroes also manifest different autonomic responses to slides portraying Negroes than do those more favorable in their verbal response? The 46 subjects, half of whom were males and half females, were divided into a prejudiced and unprejudiced group on the basis of a verbal attitude test. They were then exposed to photographic slides portraying Negroes and whites in various combinations

of race and sex. The finger pulse and GSR activity of each subject was recorded during these presentations.

The data indicate that greater GSR responses (but smaller FP responses) were given to Negro slides by prejudiced subjects, but the autonomic activity was influenced by the sex of the subject as well as the race and sex characteristics of the stimulus slides. Such autonomic activity may be considered as another dimension of attitudinal behavior to be considered along with the verbal and overt action dimensions. Additional research is needed before simple meaning can be assigned to the involvement of autonomic activity in attitudinal behavior.

REFERENCES

- ALLPORT, G. W. Attitudes. In C. Murchison. (Ed.), *Handbook of social psychology*. Worcester, Mass.: Clark Univer. Press, 1935.
- CAMPBELL, D. T. The indirect assessment of social attitudes. *Psychol. Bull.*, 1950, **47**, 15-38.
- DAVIS, R. C., SIDDONS, G. F., & STOUT, G. L. Apparatus for recording autonomic states and changes. *Amer. J. Psychol.*, 1954, **67**, 343-352.
- EDWARDS, A. L. *Experimental design in psychological research*. New York: Rinehart, 1950.
- GREEN, B. F. Attitude measurement. In G. Lindzey (Ed.), *Handbook of social psychology*. Cambridge, Mass.: Addison-Wesley, 1954.
- RANKIN, E., & CAMPBELL, D. T. Galvanic skin response to Negro and white experimenters. *J. abnorm. soc. Psychol.*, 1955, **51**, 30-33.
- WESTIE, F. R. A technique for the measurement of race attitudes. *Amer. sociol. Rev.*, 1953, **18**, 73-78.

Received January 31, 1958.

EFFECT OF CHANGES IN SELF-ESTEEM UPON IMPULSIVENESS AND DELIBERATION IN MAKING JUDGMENTS¹

JAMES E. DITTES

Yale University

SEVERAL important and frequently studied phenomena involving judgments depend for their effect upon a variable that is usually taken for granted and assumed to be constant: the occurrence of cognitive closure. All but the most trivial situations that call for judgment are characterized by ambiguity and incompleteness in the information that can be grasped initially and by the possibility of continuing assimilation of information. The state of ambiguity and the assimilation of information end when a person forms a closure, that is, when he stops deliberating and perceives a clear structure, coherence, or meaning.

Studies of conformity or rigidity generally present an individual with a piece of information and determine how closely his judgment approximates some previous judgment, either his own or another's. Variations in the readiness with which the subject forms any judgment at all, however, are generally overlooked. But without a closure, the apparent rigidity or conformity could not occur, and there exists the strong possibility that the two variables are frequently confounded. A finding of strong conformity may mean no more than that the person is motivated to form a closure impulsively and that he imposes on the material the most easily accessible structure or meaning, namely that provided by another.

Similarly, other phenomena, such as field dependence, suggestibility, ethnocentrism, resistance to free association in psychotherapy, several types of intellectual inefficiency, and some characteristics of obsessive behavior may frequently be confounded with the variable of closure and may be related as members of the "closure family." The possibility appears great enough to warrant an attempt to study the closure phenomenon itself and to discover some of its determinants, preliminary to an

attempt to unravel it from "pure" conformity, "pure" rigidity, etc.

Several lines of research have called attention to impulsiveness of closure and have attributed it to "effort after meaning" (Bartlett, 1932) or "intolerance of ambiguity" (Cohen, Stotland, & Wolfe, 1955; Frenkel-Brunswick, 1949; Smock, 1955). Such a conceptualization tends to isolate one extreme of a variable and to regard it as a discrete characteristic, fairly stable within any individual and diagnostic of particular personality characteristics. Closure may be impulsive or delayed, but in every judgment it occurs. It is proposed here to regard impulsiveness or delay as representing differing strengths of a general response tendency to achieve closure. These strengths can be measured by differing latencies in situations that also arouse motives that oppose the closure tendency—e.g., the desire for a qualitatively superior judgment. A strong response tendency produces closure at a lower threshold—i.e., with less information assimilated—than does a weaker response tendency.

Two determinants of the strength of closure are considered in this study. One is based on the assumption that the act of closure commonly acquires a reward value as a means of enhancing self-esteem. Self-esteem is regarded as a general sense of self-assurance or of adequacy, depending on a variety of internal and external stimuli (including the approval of others, achievements, and reassuring self-verbalizations), all of which are designated here as sources of self-esteem. Any damage to self-esteem, it may be hypothesized, increases the strength of the tendency to achieve closure as one means of restoring or enhancing self-esteem. The other determinant is a personality characteristic designated as general desire for clarity.

The assumptions and hypotheses guiding this research may now be stated more explicitly.

Assumption 1. The relation between self-

¹ Based on a doctoral dissertation at Yale University. Much appreciation is due members of the committee, Arthur R. Cohen, George F. Mahl, and Leonard W. Doob, Chairman. Robert Hyatt rendered valuable assistance in conducting the experimental sessions.

esteem and its sources is a negatively accelerating increasing function, approaching asymptote (see Fig. 1). If self-esteem is low, a single source provides a relatively great increment. If self-esteem is already high, the same source provides a relatively small increment.

Assumption 2. The desire to increase self-esteem functions as a drive, the strength of which may be represented by the difference between the existing amount of self-esteem and the projected maximum. Various strengths of such a drive are represented in Fig. 1 as vertical distances between the curve and the level of asymptote. This desire to increase self-esteem motivates the seeking and acquisition of additional sources. Which particular sources a person seeks depends on (a) the type of sources that have acquired self-esteem value for him and which he has been reinforced in seeking and (b) opportunity afforded by existing circumstances.

Assumption 3. The act of achieving closure is one commonly acquired source of self-esteem. Persons feel more self-assurance when they have a cognitive mastery of a situation. This reward value of closure is the result of previous experience in a wide variety of circumstances in which important need satisfactions are dependent upon cognitive closure. Such experiences may be as varied as the child's learning to name objects for the sake of parents' approval to the reduction of tension that follows the labeling or clarification of emotions in psychotherapy. They include all the instances in which action leading to the gratification of needs, even as basic as hunger, is dependent on a previous meaningful appraisal of the stimulus field. Individual and cultural differences in the amount of reward value that closure carries are to be expected, depending upon differences in learning experiences due to such factors as variation in parents' tendency to reward cognitive achievements. But the wide adaptive value of closure probably results in the acquisition of a generally high reward value in any culture.

This theoretical analysis makes the simpler assumption that closure is a source and that it provides an increment of self-esteem. The actual facts are probably more complicated. The presence of ambiguous, unstructured

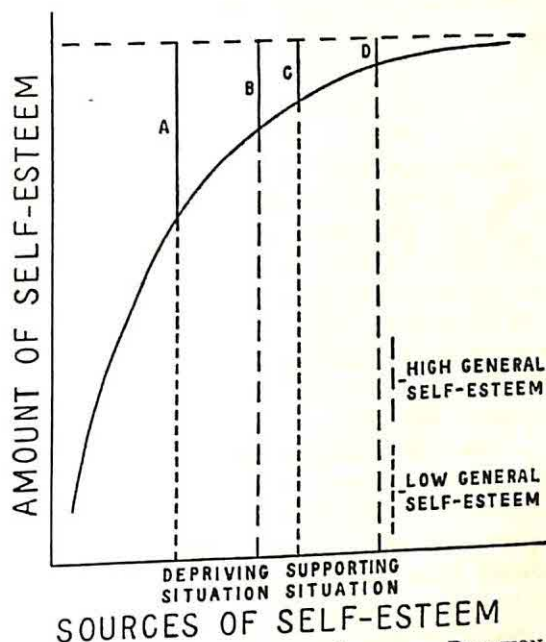


FIG. 1. REPRESENTATION OF ASSUMED RELATION BETWEEN SOURCES AND AMOUNT OF SELF-ESTEEM

(Assumed strength of motivation to increase self-esteem is represented by the distance between the curve and the level of asymptote. Representative strengths are indicated by lettered lines for persons with high and low levels of initial self-esteem in depriving and supporting situations.)

stimuli undoubtedly acquires threatening or anxiety-arousing properties and, in effect, produces a decrement in self-esteem. The derivation of the hypotheses would not be substantially affected if this assumption were made instead of, or in addition to, the assumption that closure increases self-esteem.

Assumption 4. Among other determinants of the strength of the closure tendency, there seems to be a set of persistent and general behavioral tendencies that may be designated as a *desire for clarity*. Such a trait may be indicated by the desire for explicit goals and purposes, for routine in daily tasks, for precision of thought, and for unambiguity in interpersonal relations. The desire for clarity may be regarded as a general "life-style" with diverse origins and functions, not considered here, and firmly enough entrenched to operate relatively autonomously of any situational variation. The strength of the closure tendency is assumed to be a direct function of the strength of the desire for clarity.

Hypothesis I: Persons whose self-esteem is damaged or threatened by the withdrawal or deprivation of a source of self-esteem are likely to show more impulsive closure than persons whose self-esteem is relatively guaranteed or enhanced by a similar source. This hypothesis follows directly from Assumptions 1, 2, and 3. Deprivation increases drive to enhance self-esteem and hence increases striving for the acquisition of any available sources, including closure. The increased striving for closure, in effect, establishes a lower threshold, in terms of amount of information, at which closure is achieved.

Hypothesis II: Among persons experiencing a situationally induced loss of self-esteem, those with lower general self-esteem show more impulsive closure than those with higher general self-esteem. Aside from any single situational influence, persons probably maintain a relatively even, persistent level of self-esteem, which is related to impulsiveness of closure independently of situational differences. This relation is more likely to be apparent among persons in a situationally deprived condition than among those whose self-esteem is enhanced in the situation. When the situation is supporting, all self-esteem is pushed more or less uniformly high—on the flat part of the curve in Fig. 1. When the situation is depriving, however, the more persistent personality differences are effective. This relation is illustrated by the greater difference between Lines *a* and *b* than between Lines *c* and *d* in Fig. 1.

Hypothesis III: The difference predicted by Hypothesis I is greater among persons with generally low self-esteem than among persons with generally high self-esteem. If general self-esteem is high, the addition or withdrawal of a single source within the situation has relatively little effect. If it is low, the variation of a single source has greater affect. In Fig. 1, the difference between Lines *a* and *c* is greater than the difference between Lines *b* and *d*.

Hypothesis IV: Among persons not experiencing a situational deprivation of self-esteem, those with greater general desire for clarity show more impulsive closure than those with less desire for clarity. This prediction follows directly from Assumption 4. Even in the absence of situational conditions motivating impulsiveness of closure, a strong general

desire for clarity may provide the motivation. However, when there is a strong motivational arousal of impulsiveness, as hypothesized for the condition of deprived self-esteem, the correlation between impulsiveness of closure and desire for clarity is likely to be obscured.

Hypothesis V: The difference predicted by Hypothesis I is greater among persons with lower general desire for clarity than among persons with high desire for clarity. If Hypothesis IV holds, it necessarily follows that the situational difference tends to be effective only among persons with low desire for clarity.

The last four hypotheses, concerning the interaction between the situational conditions and the two personality characteristics, self-esteem and desire for clarity, may be summarized as follows: When either personality characteristic is weak, the situation difference has more effect, since then a person is more vulnerable to situational variation (Hypotheses III and V). When the situation supplies strong stimulation that has the same effect as the personality characteristic—i.e., when it supports self-esteem or when it provides incentive to closure—the effects of differences in the personality characteristic are obscured. When the situation is negative with respect to the personality characteristic—i.e., when there is deprivation of self-esteem sources or when there is an absence of incentive to closure—the personality difference has greater effect (Hypotheses II and IV).

METHOD

Subjects

One hundred and four volunteers from the Yale freshman class met in 20 small groups of up to six members each for what they believed was a session intended to demonstrate good discussion techniques. Experimental procedures were replicated in each group.

Ss' motivation was heightened by offering a prize of \$60 and prestigious recognition to the discussion group performing the best, by letting the Ss believe that the discussion and task materials assigned their group were chosen to meet topic preferences that they had expressed during recruiting, and by stressing the importance of the research. A tape recording was made of each group's discussion to emphasize the impression that their discussion would be studied later by social scientists.

Manipulation of Self-Esteem

The first half of each session was devoted to a discussion of juvenile delinquency, based on case material

presented to the group. This discussion was interrupted three times to allow members to rate each other anonymously on the following scale:

How desirable is it that this person be kept in the group?

—extremely desirable

—very desirable

—somewhat desirable

—not very desirable, but he should be kept in

—he should be rejected from the group.

As a rationale for the request for ratings, it was explained that the investigator was interested in knowing how individual members appeared to adjust to such a situation, and that such ratings would also enable the group to improve its performance by eliminating any members who appeared detrimental to its success. *E* thumbed through the ratings as they were made and discarded them in a wastebasket.

During an intermission after the discussion of delinquency, *E* turned the conversation to these ratings and, after eliciting *Ss*' interest in seeing them, he retrieved the slips from the wastebasket and distributed them. Each *S* was thus allowed to see privately what he thought were the ratings made of him by the other members. These were, in fact, fictitious ratings prepared in advance and substituted for the originals in the wastebasket. No *S* ever indicated doubt of the genuineness of the ratings.

So that *Ss* would similarly interpret the ratings, *E* announced that he had noticed the average ratings to be around "very desirable." In each group, some *Ss* found their ratings clustered at "extremely desirable," some at "very desirable," and some at "somewhat desirable." The conditions were assigned at random around the discussion table and varied for each session.

The three conditions were intended to induce feelings of high, average, and low acceptance. Evidence of the *Ss*' direct reactions to the conditions was obtained on a questionnaire at the conclusion of the session and is reported in Table 1. *Ss* in the high and average conditions did not differ substantially in their reactions. They both reported being pleased rather than disturbed by the ratings, and on the other measures they showed similar reactions. Consequently, the data for these conditions are pooled in reporting results and regarded as a single condition of high acceptance. As compared with the high and average conditions, *Ss* in the low condition felt disturbed, less accepted, and less adequate; and they engaged in the defensive maneuver of devaluing their membership in the group.

A second type of threat to self-esteem was also introduced. At each of the levels of acceptance, half the *Ss* were given the impression that the evaluation by their peers was stable and unlikely to change, and half believed that the evaluation was conditional and very likely to change. An unstable rating was expected to threaten self-esteem and produce the same closure behavior as predicted for low acceptance. Direct measures indicated that this manipulation was less clearly disturbing to *Ss*, and it produced less clear differences in closure. Results among *Ss* who felt threatened by unstable ratings generally achieved statistical significance in the predicted direction. These are reported in

TABLE 1
DIRECT EFFECTS OF EXPERIMENTAL CONDITIONS

Measure ^a	Experimental Conditions			<i>p</i> Value High & Average vs. Low
	High	Average	Low	
Recall of received ratings	3.5	2.6	1.8	< .01
Pleased or disturbed	4.0	2.0	-.8	< .001
Sense of adequacy among peers	4.5	4.7	3.8	< .02
Valuation of group membership	5.0	5.2	4.3	< .005

^a Typical questionnaire items for each measure:

Recall: "How do you think the other members of the group rated you?"

Pleased or disturbed: "Were you pleased or disturbed when you saw the ratings?"

Sense of adequacy: "In other groups you've been in, how much do others generally like working with you and having you in the group?"

Valuation of membership: "If this group should be invited to meet again or to serve as a demonstration panel, would you want to continue working with it?"

Dittes (1958). In this report, the stable and unstable conditions are simply combined at each level.

Measures of Impulsive Closure

Parable Task. *Ss* were presented with a one-paragraph story, written in Biblical idiom and represented as a portion of a recently discovered scroll, and asked to write their opinions concerning the meaning of the parable. The account was essentially incoherent and unstructured. However, it contained a large number of familiar religious symbols that could, with some selection and distortion, be forced to yield a coherent meaning, though no two *Ss* ever found the same meaning. Subsequently, two judges decided, on a three-point scale, whether the response was a positive statement of meaning or an indication of uncertainty as to meaning. As for all the tasks, the scoring categories were defined in advance of the experimental sessions, and the judges worked independently, without knowledge of the *Ss*' performance on other tasks or of the experimental conditions to which the *Ss* had been assigned. The categories were so defined that the higher score represented more impulsive closure. The judges showed absolute agreement on this task in 86% of the judgments. The contingency coefficient representing the agreement in sorting into the three categories was .75.

Impression Formation Task. *Ss* were given two lists of traits that were represented as words used by close friends to describe two students; they were instructed to write their impression of each man. Each list contained a dominant set of traits and a subordinate set which was incompatible with the dominant set. In one instance, for example, the list was: individualist, suspicious, belligerent, considerate, self-concerned, thoughtful, stubborn, tolerant. The two lists were balanced so that dominant traits in one were favorable, in the other

unfavorable. Subsequently, two judges decided, on a three-point scale, the extent to which the subordinate traits were integrated into the impression. A statement based exclusively on the dominant traits was regarded as the most impulsive. There was absolute agreement on 72% of the judgments, and the agreement yielded a contingency coefficient of .64.

Jammed Message Task. Ss listened to a sound tape on which a phrase was repeated 20 times with decreasing volume of masking noise on each repetition. They indicated the number of the repetition on which they were "quite sure" that they recognized the phrase by selecting the correct phrase from among five alternatives read before the tape was played. The complete task comprised three such phrases. The measure of impulsiveness was the repetition on which the S first claimed to recognize the phrase, this summed for the three phrases. Because of significant differences between groups, the deviation of each S from the mean of his group was taken as his score, with a positive sign given to the earlier response.

Summary of Task Characteristics. Each task presented the S with unstructured, ambiguous elements that he regarded as potentially coherent and meaningful. Each task presented cues that provided the opportunity for an immediate structuring; for the three tasks these were, respectively, commonly used religious symbols, a quickly apprehended set of dominant traits, and fragments of a spoken phrase. Each task also provided cues prompting delay and withholding of early closure. These were of two types: (a) information that was not readily assimilable and therefore had to be ignored or distorted if a quick closure was to be achieved; and (b) indications that the available information was incomplete and that more was forthcoming. The first type was represented in the parable and impression formation tasks by the presence of elements incompatible with the kind of judgment that could be formed impulsively. The second type was represented in the parable by the announcement that it was but a portion of a translation and in the jammed message task by the knowledge that the phrase on the tape was about to be heard more clearly in another 15 seconds.

Each task determined individual differences in threshold of closure. The parable task presented a constant stimulus situation, approximately at the average threshold and determined for each S whether he formed a closure. The other two tasks presented gradually assimilable information and determined how much information each S assimilated before he formed a closure.

Combined Index. Since all the tasks were generated from the same rationale to measure a single behavioral tendency, the scores for each task were combined to form a more reliable over-all index of closure strength for each S by summing the scores after weighting them by factors to equalize their variance. The low though consistently positive correlation among the task scores (ranging from .06 to .19) is presumably due to the fact that each task elicits peculiar skills, interests, or background that override the experimental effects for individual Ss, but still leave important group differences.

Personality Measures

General Desire for Clarity. Six items on a questionnaire before the experimental session sampled the Ss' desire for clarity in a variety of life situations. Some items were taken from a scale used by O'Connor (1952) and some from a questionnaire used by Cohen et al. (1955). The items were as follows:

1. It is always better to have a definite course of action rather than to be vacillating among several possibilities.

2. A smart person gets his life into a routine so that he is not always being bothered by petty details.

3. In writing a theme or essay, clear organization of the material is of utmost importance.

4. A committee works best when the responsibilities of each member are clear.

5. It is better to keep on with the present method of doing things than to take a way that might lead to chaos.

6. Textbooks which are unclear are really useless, no matter how complete or otherwise valuable they may seem to be.

Ss indicated agreement or disagreement with each item on a six-point scale. Reliability is indicated by a product-moment correlation, corrected for length, of .84 between the scores of the odd- and even-numbered items.

For purposes of comparison, the distribution of scores was divided into quarters at the most natural cut-off points, before any of the results were examined. No alternative division was undertaken. The self-esteem scores were divided into thirds on the same basis; since these scores tended to be tri-modal, to have separated them into quarters would have arbitrarily divided clusters of scores.

Self-esteem. General self-esteem was measured with an index made up of three components. (a) Eleven self-rating items were used on the questionnaire before the experiment. Most of these were adapted from Murray (1938) as was the form of administration. The S indicated on a six-point scale whether each of the items applied more or less to him than to the average man his age. Typical items were:

TABLE 2
AVERAGE CLOSURE SCORES

Task	Experimental Conditions			p Value High & Average vs. Low
	High	Average	Low	
Parable	2.0	2.0	2.5	< .05 ^a
Impression formation	2.1	2.2	2.3	< .35
Jammed message	-1.2	-.7	1.9	< .05
Combined index	43.6	46.5	54.5	< .002

^a Based on a chi-square test of the distribution among the three scoring categories, because scores were not normally distributed. Other tests are *t* tests.

I feel embarrassed and uncomfortable in the presence of people who are socially gifted.

I feel anxious and uncertain when I am suddenly faced by a critical situation.

The product-moment correlation between the scores on the odd- and even-numbered items, corrected for length, was .82.

(b) A measure of sense of adequacy among peers was obtained with four items on the questionnaire at the end of the session. This score was found to be significantly correlated ($r = .34, p < .01$) with the before-session measure of self-esteem, after differences attributable to experimental manipulation were removed by taking as each S's score his deviation from the mean of his experimental condition.

(c) The actual ratings that Ss received from each other, though never actually saw, were justified as another indication of level of self-esteem by the assumption that because of the brief period of acquaintance preceding the actual ratings, these were likely to be based on behavior (e.g., active participation vs. reticence) which is largely a function of a person's own appraisal of himself. This measure correlated significantly ($r = .25, p = .01$) with self-ratings.

These three measures had almost identical variances and were summed to provide the combined index. The decision to use this combined index to measure self-esteem was made before the results were examined.

RESULTS

Hypothesis I. Persons made to feel well accepted consistently showed less impulsiveness than those made to feel poorly accepted. Table 2 reports the results for each of the three tasks and for the combined index, which showed a highly significant difference ($t = 3.32, p < .002$)² between the high and average

² Throughout, all probability values are derived from using both tails of the statistical distribution.

and the low conditions. The intended high and average conditions, which showed similar direct reactions to the experimental manipulation as reported above, also produced substantially similar results, though tending consistently to differ in the predicted direction.

Results relating to the hypothesized interaction of the experimental conditions with the personality characteristics are reported for the combined index of closure in Tables 3 and 4.

Hypothesis II is tentatively supported by the second column in Table 3. Within the condition of low acceptance, the third of the Ss with lowest self-esteem were more impulsive

TABLE 3
AVERAGE SCORES ON COMBINED INDEX OF CLOSURE
AT DIFFERENT LEVELS OF SELF-ESTEEM

Level of Self-Esteem	High Acceptance	Low Acceptance	Difference	p Value
High	47.3	52.6	5.3	< .30
Average	42.3	52.1	9.8	< .05
Low	45.8	58.8	13	< .05

TABLE 4
AVERAGE SCORES ON COMBINED INDEX OF CLOSURE
AT DIFFERENT LEVELS OF DESIRE FOR CLARITY

Level of Desire for Clarity	High Acceptance	Low Acceptance	Difference	p Value
High	49.2	53.8	4.6	< .40
High average	44.2	51.9	7.7	< .20
Low average	44.4	53.8	9.4	< .10
Low	43.5	58.0	14.5	< .01

TABLE 5
AVERAGE CLOSURE SCORES ON SEPARATE TASKS FOR DIFFERENT LEVELS OF SELF-ESTEEM AND
DESIRE FOR CLARITY

	Parable Task				Impression Formation Task				Jammed Message Task			
	High Acc.	Low Acc.	Diff.	p	High Acc.	Low Acc.	Diff.	p	High Acc.	Low Acc.	Diff.	p
Self-esteem:												
High												
Average	1.1	1.3	.2		2.2	2.2	0		-.5	3.6	4.1	
Low	.8	1.6	.8	< .05	2.1	2.2	.1		-1.5	.2	1.7	
Desire for Clarity:												
High	.9	1.7	.8	.10	2.2	2.5	.3	.10	-.4	2.0	2.4	
High average									.4	-.1	1.6	
Low average	1.3	1.5	.2		2.2	2.4	.2		-3.2	-1.6	2.3	
Low	.8	1.5	.7		2.4	2.4	0		1.8	4.1	7.2	.02
	.8	1.6	.8	< .05	2.1	2.0	-.1		-2.4	4.8		
	1.0	1.6	.6		2.0	2.2	.2					

($t = 1.7, p < .10$) than the upper two-thirds. Among Ss in the high condition, no variation approached significance. *Hypothesis III* is supported by the trend of a progressively greater difference between experimental conditions in lower thirds of self-esteem.

As predicted by *Hypothesis IV*, persons with greatest desire for clarity tended to be impulsive even when they were in the experimental condition of high acceptance. Within the high condition, Ss in the top quarter of desire for clarity were more impulsive ($t = 2.08, p < .05$) than those in the lower three-quarters. *Hypothesis V* is supported by the tendency for the difference between experimental conditions to become progressively greater with less desire for clarity.

Results with the separate tasks, reported in Table 5, were substantially like those with the combined index, although significance levels are lower as a result of the greater unreliability of any single estimate of closure.

DISCUSSION

Low evaluation by peers clearly results in greater impulsiveness of closure on various types of conceptual tasks. This effect depends in an important way, however, upon certain entrenched personality characteristics. The findings can be accounted for in terms of the damage to self-esteem produced by low acceptance and of the capacity to enhance self-esteem that closure commonly acquires. Impulsive closure becomes a compensating source of self-esteem.

There are various alternative accounts that can be advanced, although none of them seems congruent with all the available findings. One important possibility is that poorly accepted persons produced impulsive closure in order to regain acceptance or to forestall rejection, that is, for its direct *instrumental* value. The mediation of changes in self-esteem, it may be argued, need not be invoked. Such an explanation, however, requires certain assumptions and predictions that are not supported by the available data:

1. It must be assumed that Ss experienced some direct association between impulsiveness of closure and acceptance by these peers. If Ss were trying to regain acceptance by the group through impulsive closure, they must have

supposed that the group (or *E*) would reward impulsiveness and punish deliberation. However, because this possibility raises an important problem in interpreting some previous similar findings (e.g., Smock, 1955), care was taken in designing the procedures to insure that impulsiveness could not become an implicit norm within these groups any more than delayed, more cautious closure. The only instructions were general suggestions to Ss to engage in good discussion and to make good decisions, and they were left to interpret "good" in any manner they chose. The last two of the tasks were presented as incidental exercises, completely unrelated to even this group goal of good discussion, and the Ss were clearly instructed that their performance was not to be known by the other members.

On the contrary, it seems more likely that the groups developed a norm for deliberative judgments that tended to work against the actually obtained results. The academic environment from which the Ss were drawn and in which the experiment was held tends to reward quality of closure rather than impulsiveness on complex judgments of the sort required here. Observation of the discussion and some questionnaire responses indicate that within the experimental sessions Ss did feel pressure toward delayed closure. These are reported elsewhere (Dittes, 1958). None of the evidence suggests that Ss consciously experienced any pressures toward impulsiveness.

Although time pressures may have been experienced occasionally by Ss, impulsiveness of closure is not necessarily related to response time; rather, it refers to whether the thought process occurs in an essentially structured ambiguous framework. A person might spend much time and effort interpreting a parable, for example, by compiling meanings of the different symbols and perhaps even by shifting from one interpretation to another. But the impulsiveness exists in the primary effort of grasping for some certain meaning. Neither of two possible indications of response time, the rank orders within each group of beginning and of completing a response to the parable, was correlated with scores of impulsiveness on that task.

2. Since it is well established (Back, 1951; Dittes & Kelley, 1956; Festinger, 1950) that

efforts to regain group acceptance, such as this explanation supposes, are in part a function of Ss' evaluation of their membership in the group, this alternative interpretation must predict that the effects would be greater among Ss who most highly valued their membership than among those who least valued their membership. However, when Ss are divided according to their indication on the final questionnaire of high or low evaluation of their membership in the group, virtually identical results are obtained with those who set a high value on membership as with those who set a low one. The respective scores of these two groups on the combined index under conditions of high and low acceptance were 45 and 55, and 46 and 55. The valuation measure must be regarded as reliable because, in other results obtained in the experiment and reported elsewhere (Dittes, 1958), it was shown to be useful in predicting results that could be clearly attributed to a group conformity effect.

The above discussion concerns the possible interpretation that Ss performed impulsive closure because they believed that *impulsiveness* would be rewarded by greater acceptance. Another possibility is that *closure* has reward value, not as a source of self-esteem as posited by Assumption 3, but exclusively by association with the approval of other persons. A striving to improve acceptance, rather than a striving to improve self-esteem, may mediate the results; but if this is so, it must be regarded, for the reasons indicated just above, as functioning exclusively on the basis of learned associations and independently of any realistic expectations about potential approval or disapproval within the experimental situation.

Such a striving to improve acceptance would be functionally equivalent with striving to improve self-esteem, and if self-esteem were considered as based exclusively on the approval of others, the two assumed motives would be identical conceptually. The self-esteem hypothesis proposes that the self-esteem value of closure is based on other need gratifications as well as on praise by others and therefore that the results of this experiment would be substantially the same if self-esteem were manipulated by varying some other source, such as achievement and failure experiences.

A third possible explanation would attribute

the effects to the distraction caused by the treatment of low acceptance. It is possible that poor acceptance elicits aggressive, depressive, or other reactions within Ss which could have the effect of distracting them from full attention or motivation on the subsequent tasks. Lessened attention on the tasks, it might further be argued, would lead to off-hand, less careful responses, which would be scored as impulsiveness. Asch (1952), for example, has proposed that when attention is directed toward ego-concerns (as the condition of low acceptance may do) it is correspondingly withdrawn from the task, and this explanation has been used by Shipley (1956) to account for findings similar to those reported here. A related theory of "interference" is used by Moffitt and Stagner (1956) to account for similar findings.

There is no direct evidence available on the extent to which Ss were so distracted. But related data suggest that distraction effects were not critically involved: In the first place, any such distraction is likely to be reflected in responses to the items measuring Ss' motivation to be in the experimental group. One of the items, for example, asked how hard Ss tried to perform well. Yet, as already reported, results are virtually identical among Ss who scored low on the measure of motivation and those who scored high. Secondly, examination of the definitions of the scoring categories for the tasks does not suggest that behavior scored as impulsive is likely to be the result of casual, off-hand response nor of inattentive, inefficient, or primitive intellectual functioning, except possibly on the impression formation task. Early attempts to decipher the phrase on the jammed message task and the intellectual agility required to discover some meaning in the parable suggest greater involvement and attention, not less.

SUMMARY

Closure is defined as the structuring of otherwise ambiguous stimuli so as to provide meaning and to end deliberation. Assuming that (a) such closure commonly acquires reward value as a source of self-esteem through learning experiences in which it has been associated with achievements, praise by other persons, and other fundamental sources of self-esteem;

and (b) damage or threat to self-esteem produces a stronger motivation to increase self-esteem by various means, including closure; it was hypothesized that a situationally induced threat or injury to self-esteem would result in greater impulsiveness of closure in making complex judgments. Such impulsive closure may be the equivalent of behavior sometimes attributed to "effort after meaning" or "intolerance of ambiguity," and may be an important component of rigidity, conformity, suggestibility, field dependence, ethnocentrism and similar members of a possible "closure family."

Self-esteem was manipulated by communicating to Ss, meeting in small groups, fictitious information as to how well each one was regarded by the others in the group. Impulsiveness or delay in closure was determined by three different behavioral measures. The threat to self-esteem of being poorly accepted resulted in significantly more impulsive closure as represented by tendencies (a) to find a positive meaning in an essentially incoherent prose passage rather than to acknowledge its incoherence, (b) to base an impression of another person exclusively on more prominent traits and to ignore inconsistent traits, and (c) to claim to recognize a disguised phrase, which was repeated on a sound tape with ever-increasing clarity, earlier in the sequence of repetitions.

Interactions were predicted between the situational conditions and personality characteristics of general self-esteem and general desire for clarity, as measured by a questionnaire. Support or threat tended to be more effective in producing differences in impulsiveness of closure among Ss with low self-esteem and among Ss with low general desire for clarity; their behavior can be regarded as less determined by persistent personality characteristics and more vulnerable to situational differences. Personality characteristics alone

tended to affect impulsiveness of closure in instances in which the situational stimulation did not obscure the personality differences: among Ss who were situationally threatened, those with low self-esteem tended to be more impulsive than those with high self-esteem, and among Ss situationally supported, there was an indication of a positive correlation between general desire for clarity and impulsive closure.

REFERENCES

- ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
- BACK, K. W. Influence through social communication. *J. abnorm. soc. Psychol.*, 1951, **46**, 9-23.
- BARTLETT, F. C. *Remembering*. Cambridge: Cambridge Univer. Press, 1932.
- COHEN, A. R., STOTLAND, E., & WOLFE, D. M. An experimental investigation of need for cognition. *J. abnorm. soc. Psychol.*, 1955, **51**, 291-294.
- DAVIDS, A. Some personality and intellectual correlates of intolerance of ambiguity. *J. abnorm. soc. Psychol.*, 1955, **51**, 415-420.
- DITTES, J. E. Effects of changes in self-esteem upon impulsiveness and deliberation in making judgments. Unpublished doctoral dissertation, Yale Univer., 1958.
- DITTES, J. E., & KELLEY, H. H. Effects of different conditions of acceptance upon conformity to group norms. *J. abnorm. soc. Psychol.*, 1956, **53**, 100-107.
- FESTINGER, L. Informal social communication. *Psychol. Rev.*, 1950, **57**, 271-282.
- FRENKEL-BRUNSWIK, ELSE. Intolerance of ambiguity as an emotional and perceptual personality variable. *J. Pers.*, 1949, **18**, 108-143.
- MOFFITT, J. W., & STAGNER, R. Perceptual rigidity and closure as functions of anxiety. *J. abnorm. soc. Psychol.*, 1956, **52**, 354-357.
- MURRAY, H. A. *Explorations in personality*. New York: Oxford Univer. Press, 1938.
- O'CONNOR, PATRICIA. Ethnocentrism, "intolerance of ambiguity," and abstract reasoning ability. *J. abnorm. soc. Psychol.*, 1952, **47**, 526-530.
- SHIPLEY, T. E. Threat to the self, the direction and breadth of attention, and the distance gradient. Unpublished doctoral dissertation, Univer. of Pennsylvania, 1956.
- SMOCK, C. D. The influence of psychological stress on the "intolerance of ambiguity." *J. abnorm. soc. Psychol.*, 1955, **50**, 177-182.

Received February 19, 1958.

STABILITY AND VALIDITY OF ACHIEVEMENT FANTASY¹

JEROME KAGAN AND HOWARD A. MOSS

Fels Research Institute at Antioch College

ALTHOUGH projective and fantasy instruments have been widely used both in clinical and research settings the reliability and validity of many fantasy variables still remains in question. This paper is the first in a series of reports dealing with the stability and validation of fantasy variables from the TAT and Rorschach and deals specifically with achievement fantasy on the TAT. Interest in achievement fantasy has been fairly intense in recent years and many studies have attempted to predict various types of achievement behaviors from fantasy responses (Atkinson: 1954, 1957; French: 1955, 1958; French & Thomas, 1958; Lazarus & Baker, 1957; McClelland & Liberman, 1949; McClelland, Atkinson, Clark, & Lowell, 1953; Vogel, Lazarus, & Baker, 1957; Winterbottom, 1952). Most workers conceive of achievement themes as a measure of the intensity of achievement motivation, and this measure has been labeled "need Achievement." However, Farber (1954) has pointed out that in an explanatory system that employs motive and response strength as independent parameters, as in behavioral theory, it may be fallacious to use fantasy behavior as a measure of motive strength. Since a fantasy response is a sample of overt behavior, it is influenced by all conditions that affect behavior, including both goal-oriented and goal-avoidant response tendencies. That is, when avoidant or inhibitory responses are associated with a motive the probability of occurrence of any motive-related response is, in part, a function of the strength of the avoidant tendencies. Thus, when one is dealing with a complex behavior like achievement, variables such as anxiety over ability to obtain an achievement goal, independent of the moti-

vation for the goal, may elicit avoidant responses that inhibit both direct achievement behavior and achievement fantasy. On these premises, no single class of responses such as fantasy themes is adequate for assessing the intensity of achievement motivation, and it seems conceptually less ambiguous to view achievement themes as a response variable, or as one index of a person's behavioral tendency to engage in achievement behaviors. This point of view seems similar to that of Atkinson (1954) who defines "need Achievement" in response terms: "... the strength of a family of perceptual and instrumental response dispositions corresponding to events in past behavioral sequences which have led to the anticipated goal state" Further discussion of the theoretical interpretation of achievement fantasy will be reserved for the latter section of the paper.

This research deals with three aspects of a fantasy measure of achievement striving: (a) its stability over time, (b) its relation to certain antecedent conditions, and (c) its prediction of overt achievement behavior.

The Fels Research Institute is well suited to undertake a study of this kind, since for a large sample of children it has continuous longitudinal data, including TAT protocols, ratings, and reports of the mother's attitudes and behavior, school grades, and continuous IQ testing. Some of the information is relevant for the above three questions. Since the TAT cards were administered every 2½ to 3 years, the long-term stability of this behavior can be assessed. Lowell (1950), using equivalent sets of fantasy stimuli and a test-retest interval of one week, found a significant correlation between two achievement fantasy scores ($\phi = +.44$). This report offers the opportunity to assess the stability of achievement fantasy over a much longer period of time. With respect to antecedent variables, it was felt that both direct, maternal concern with the child's achievement as well as the educational level of the parents might relate to the degree of achievement striving in the older child. It was assumed that if a parent valued achieve-

¹ This research was supported in part by a research grant (PHS-M-1260) from the National Institute of Mental Health of the National Institutes of Health, United States Public Health Service. The authors would like to acknowledge the fact that the ideas of L. W. Sontag concerning the relation between changes in IQ and personality variables was an important stimulus for the current study. We also wish to thank Vaughn J. Crandall for his careful reading of the manuscript.

ment behavior he would encourage and praise this behavior in the child and create reinforcement conditions that would facilitate the learning of the behavioral tendency. Thus, we anticipated a positive relation between ratings of early maternal encouragement and concern with a child's cognitive and motor development and later achievement strivings in the child. In addition, amount of formal education in the parent was selected as an indirect index of achievement concern on the part of the parent, since the acquiring of a high school diploma or college degree is generally viewed as an achievement-related goal. It is a relatively crude index, since many circumstantial factors are involved in determining educational level.

With respect to the behavioral consequences of achievement strivings, it was assumed that the desire to perform well in an intellectual test situation was one reflection of achievement strivings. In a recent monograph, Sontag, Baker, and Nelson (1958) reported that children who were competitive and initiated activities on their own showed large increases in IQ score during the childhood years. Kagan, Sontag, Nelson, and Baker (1958), using the same sample of children, found that those children who showed large childhood increases in IQ were more likely to tell TAT achievement themes than the children with decreases in IQ. It was suggested that the achievement fantasy reflected a tendency to strive for efficient performance in the IQ test situation and that this striving facilitated IQ gains. The present research also used IQ gain during childhood as an index of achievement strivings and offered the opportunity to study the relation between achievement fantasy and IQ changes where the changes were not extreme but representative of the entire distribution.

METHOD

From the files of the Institute population, 86 cases (Ss) were selected (44 boys and 42 girls) on whom the following information was available: (a) three seven-card TAT protocols with median ages of administration at 8-9, 11-6, and 14-6; (b) annual Stanford-Binet IQ scores from age 3 through 11; (c) educational background of each parent; and (d) written reports of the mother's behavior with the child based on 6 to 10 visits to the home for the first three years of the child's life. This latter information was available on only 44 cases.

Scoring of Achievement Fantasy

Each of the S's three protocols consisted of seven TAT cards (Cards 1, 5, 14, 17 BM, 3 BM, 6 BM, and 3 GF), and achievement fantasy was scored according to the scheme described by McClelland et al. (1953). Since incidence of the subcategories of the McClelland scoring system were infrequent (e.g., I+, I-, Ga+, etc.), only stories in which achievement concern or behavior was the major aspect of the plot are considered. These are scored *Ach Th* in the McClelland scheme. In addition, because there was a lack of comparability among the examiners with respect to the inquiry questions, only the spontaneous verbalization of the S was scored. Agreement between two independent raters scoring the themes was 95%.

Educational Background of the Parents

The formal educational training of each parent was assigned a score of 1 through 6 according to the following scheme: 1. eighth grade or less; 2. part high school; 3. high school graduate; 4. part college; 5. college graduate; 6. post graduate training.

Early Maternal Concern with Achievement and Development

For 44 of the 86 children (20 boys and 24 girls) there were lengthy reports, based on visits to the home, in which the interaction of mother and child was described. These reports were studied for the first three years of the child's life, and independent ratings were made by the authors for a variety of variables. All variables were rated on a seven-point scale. The variable relevant for this report was defined in terms of the degree to which the mother showed concern and anxiety over the child's cognitive and motor development and the degree to which she placed excessive expectations on the child's achievements. It reflected the degree to which the mother pushed the child's development beyond his abilities and her concern with his general achievement level. Major sources of evidence included concern over the age when the child walked, talked, rolled over, showing off of the child's skills and abilities and maternal dissatisfaction with the child's development. The interrater reliability was +.80 (product-moment correlation).

IQ Changes

Each child was given the Stanford-Binet, Forms L and M alternately, semiannually from ages 3 to 5 and annually from ages 6 through 11. All tests were administered by the same psychologist.² The mean IQ for the entire Fels population is about 120 (*SD*, 15). For each S, a smoothed plot of his IQ scores was obtained by averaging his three IQ scores around each age. For example, a child's smoothed or average IQ at age 6 was the result of averaging his IQ scores at

² The authors wish to acknowledge the assistance of Virginia L. Nelson who administered all of the IQ tests and who has been helpful during all phases of the research.

ages 5, 6, and 7; his smoothed IQ at age 10 was the average of his IQs at ages 9, 10, and 11. The procedure tends to remove the chance variation associated with any one IQ score and has been used in other studies (Kagan et al., 1958; Sontag, 1958). Each S's smoothed IQ at age 6 was then subtracted from his smoothed IQ at age 10, and the resulting differences were used as the measure of IQ change.

RESULTS

Stability of Achievement Fantasy

There was a tendency for the proportion of Ss reporting achievement themes to increase with age, and there were no marked sex differences at any one administration. Table 1 shows the percentage of each group reporting one or more achievement themes on each of the three protocols.

The average number of achievement themes per protocol was one, and Cards 1 and 17 BM, both suggestive of achievement plots, accounted for 76% of all achievement themes. Card 1 elicited more achievement themes than any card (43% of all themes) but, with age, there was an increase in the incidence of achievement themes to Card 17 and a decrease to Card 1. The typical achievement theme to Card 1 concerned a boy who wanted to master the violin and/or become a famous violinist, while the typical achievement story to Card 17 BM involved a person who was in a rope-climbing contest and wanted to do his best to win.

Since achievement fantasy was scored as either present or absent on any one protocol, the phi coefficient was used to express the reliability of the presence of the fantasy over the three administrations. Since there were no marked sex differences with respect to the reliability of the fantasy, the sexes were pooled in computing the final phi coefficients. Table 2 shows the value of phi and the associated one-tailed significance values for the intercorrela-

TABLE 1
PERCENTAGE REPORTING ACHIEVEMENT THEMES
ON THE THREE PROTOCOLS

Protocol	Boys	Girls	Boys and Girls
1			
2	40.9	35.7	38.4
3	47.7	45.3	46.5
	59.1	59.5	59.3

TABLE 2
RELIABILITY OF ACHIEVEMENT FANTASY
FOR THREE PROTOCOLS
(Phi coefficients)

	Protocol 1	Protocol 2	Protocol 3
Protocol 1	—	+.32**	+.22*
Protocol 2		—	+.16

* $P < .05$.

** $P < .01$.

TABLE 3
FORMAL EDUCATION OF THE PARENTS
(Percentage of group with parents at each
level of education)

Educational Level	Boys		Girls	
	Father	Mother	Father	Mother
1. Eighth grade or less	18.2	11.4	33.3	11.9
2. Some high school	9.1	13.6	9.5	11.9
3. High school graduate	20.5	31.8	16.7	26.2
4. Some college	6.8	22.7	21.4	26.2
5. College graduate	22.7	20.5	11.9	23.8
6. Post graduate	22.7	0.0	7.2	0.0

tions among the three protocols. The occurrence of achievement fantasy on Protocol 1 was significantly associated with occurrence of the fantasy on Protocols 2 and 3.

Educational Level of the Parents

In many families, one parent had more formal education than the other, and since the values of the same-sex parent were assumed to have special influence on the identification process in the child, analyses of this variable were kept separate for boys and girls. Table 3 summarizes the education of the 86 parent pairs under study.

The relation between educational level of the parents and occurrence of achievement fantasy on Protocol 1 was assessed by dichotomizing the variable of educational level into "Some college training" versus "No college training" and relating this dichotomy to presence or absence of achievement fantasy on the first protocol. Table 4 shows the resulting phi coefficients for boys and girls.

Children from homes in which both parents had college training were not much more apt to report achievement themes on Protocol 1 than children from homes in which neither parent had college training. The mean parental

TABLE 4
RELATION BETWEEN ACHIEVEMENT
FANTASY ON PROTOCOL 1 AND
EDUCATIONAL LEVEL OF
PARENTS
(Phi coefficients)

	Education of Both Parents	Education of Father	Education of Mother
Boys	+ .17	+ .28*	+ .11
Girls	+ .09	- .11	+ .01

* $P < .05$.

educational level of the group reporting achievement themes was 3.5, while the mean for the group without achievement fantasy was 3.2. Thus, achievement fantasy on the first protocol was not significantly related to the educational background of the family as a whole. However, for boys, there was a positive relation between education of the father and occurrence of achievement fantasy. The mean educational level of the fathers of boys with achievement themes was 4.7; the mean for the fathers of boys without achievement fantasy was 3.1 ($t = 3.20$; $p < .01$). In order to assess the stability of these results, phi coefficients were computed relating achievement themes on Protocols 2 and 3 with both the educational level of each parent and the family as a whole. For Protocol 2 there was no relationship between occurrence of achievement fantasy and the education of either or both parents. For Protocol 3, there was a positive relation, for girls only, between occurrence of achievement themes and maternal level of education ($\phi = +.26$; $p < .05$).

In general, for this middle-class sample, level of parental education did not show any consistent or highly significant relationship to achievement fantasy in the child. However, the positive relation between achievement themes and educational level of the same-sex parent, which occurred for boys on Protocol 1 and for girls on Protocol 3, tentatively suggests that achievement concern on the part of the same-sex parent may have a special influence on the child's achievement strivings. This statement assumes that a highly educated parent would be more concerned with achievement in his child than a poorly educated parent.

Maternal Concern with Achievement

The mean ratings on maternal concern with achievement during the child's first three years were compared for Ss with achievement fantasy present or absent on the first protocol. The mothers of girls who reported an achievement story were rated as more concerned with achievement than mothers of girls who did not report an achievement theme ($t = 2.65$; $p < .01$). The difference for the boys was negligible ($t = .79$; $p > .20$). The stability of this finding was assessed for Protocols 2 and 3, and similar results were obtained. For both Protocols 2 and 3, the mothers of the girls reporting an achievement theme were rated as more concerned with achievement than the mothers of the girls who did not report achievement fantasy ($p = .07$, $.05$ respectively). The differences for the boys were again negligible. Thus, the relation between early maternal concern with achievement and achievement fantasy in girls was relatively stable over a six-year period.

IQ Change and Achievement Fantasy

The dependent variable correlated with achievement fantasy was amount of increase in IQ over the ages 6-10. It should be noted that because of familiarity with the tests and the testing procedure, the average gain for a large sample of Fels children over this four-year period is 6 points and either no change or a loss in IQ is atypical for the average Fels S. Table 5 summarizes the data on base level IQ at age 6, mean IQ change and range of change for children high and low in achievement fantasy on the first protocol. There was no significant difference in base-level IQ at age 6 between the sexes, or between Ss with achievement fantasy present or absent. However, Ss who told an achievement theme on the first protocol showed a significantly larger gain in IQ (+10.7 points) than those who did not tell an achievement theme (+ 5.5 points) ($t = 2.63$; $p < .01$). In order to guarantee that this result was not a function of differences in IQ at age 6, an analysis of covariance was also applied to the data. The results revealed that IQ gain was significantly associated with the presence of achievement fantasy when variation due to base-level IQ at age 6 was partialled out ($F = 5.56$; $p < .01$).

TABLE 5
RELATION BETWEEN IQ GAIN AND ACHIEVEMENT
FANTASY ON PROTOCOL 1

	Achievement Present			Achievement Absent		
	Boys	Girls	Both Sexes	Boys	Girls	Both Sexes
Mean IQ at age 6	120.2	116.2	118.4	110.2	115.4	112.9
Mean increase in IQ	11.6	9.7	10.7	7.4	3.7	5.5
Range of IQ changes	-12 to +40	+1 to +24		-11 to +26	-11 to +16	

Figure 1 illustrates the relation between the amount of IQ gain and the number of protocols in which one or more achievement themes occurred (achievement positive protocols). Evidently the relationship is not linear. Although Ss with three achievement positive protocols showed a greater gain in IQ than Ss having no achievement positive protocols ($t = 2.74$; $p < .01$), the Ss with two positive protocols showed less IQ gain than those with one positive protocol ($t = 2.05$; $p < .05$). The drop in IQ gain for Ss with two achievement protocols was more marked for boys than for girls for the difference in IQ gain between Ss with one versus two positive protocols was 9.5 points for the boys and only 2.6 points for girls.

If one regards number of achievement positive protocols as an index of the strength of achievement strivings, then this result is in agreement with other research which has shown a nonlinear relation between achievement themes and a class of overt achievement behavior (Clark, Teevan, & Ricciuti, 1956; McClelland et al., 1953). McClelland has hypothesized that Ss who show a moderate amount of achievement fantasy are characterized by both a strong desire for achievement goals and a strong fear of failure and assumes that the fear of failure elicits responses that oppose and block successful achievement behavior. In an attempt to ascertain whether fear of failure fantasy was related to amount of IQ change the mean IQ gain of the 25 Ss who reported one or more fear of failure themes on any one of the protocols was compared with the mean gain of the 43 Ss who only reported positive, success achievement themes. The mean gains were 9.7 and 9.4 points respectively indicating no marked difference between these

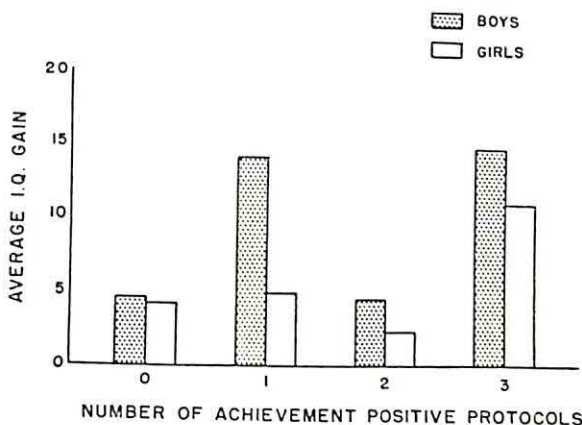


FIG. 1. RELATION BETWEEN IQ GAIN AND NUMBER OF ACHIEVEMENT POSITIVE PROTOCOLS

two groups. Thus, for this sample, fear of failure fantasy was not related to inferior performance on the IQ increase measure.

Another possibility is that the testing situation was not a strong incentive for all the Ss, and only those children who sought gratification of achievement goals in an intellectual area would strive to perform well on the intelligence test. This hypothesis was supported by information on the elementary school grades of 35 boys and 37 girls from the original sample of 86. Each child's average grade over the four-year period covering elementary grades 2-5 was computed. Then the boys and girls were ranked separately with relation to their mean grade and each distribution dichotomized at the median into high and low grade groups. Table 6 shows the distribution of achievement positive protocols for boys and girls above or below the median in school grades.

For girls there was no difference between the high and low grade groups with respect to the distribution or number of achievement-positive protocols. However, the boys with two achievement-positive protocols were more likely to have low grades than those with one such protocol ($p < .01$; exact test). This finding may help to clarify the meaning of the minimal IQ gain shown by the boys with two achievement-positive protocols. That these same boys performed poorly in school suggest that an intellectual task was not a strong incentive for them. It is tentatively suggested that their achievement strivings were directed toward goals related to nonintellectual activities like athletics or manual skills.

TABLE 6

DISTRIBUTION OF ACHIEVEMENT POSITIVE PROTOCOLS
FOR BOYS AND GIRLS ABOVE AND
BELOW THE MEDIAN IN
SCHOOL GRADES
(Percentages)

No. of Achievement Positive Pro- tocols	Boys		Girls	
	High Grades	Low Grades	High Grades	Low Grades
0	23.5	22.2	15.0	23.5
1	47.0	16.7	35.0	35.3
2	6.0	50.0	20.0	17.7
3	23.5	11.1	30.0	23.5
Total	100.0	100.0	100.0	100.0

These results seem to indicate that gain in IQ during childhood is one possible measure of achievement-strivings with respect to intellectual activity. The product-moment correlations between average school grade, for the grades 2-5, and IQ gain was $+.48$ for girls and $+.25$ for boys ($p < .01$, $< .10$ respectively). In addition, amount of IQ gain does not seem dependent on initial intellectual capacity, for the product-moment correlation between Binet IQ at age 3 and amount of IQ change was nonsignificant ($r = +.15$), while the correlation between IQ at age 3 and at age 6 was $+.73$ ($p < .01$).

Maternal Concern with Achievement and IQ Change

Since early maternal concern with achievement was positively correlated with achievement fantasy for girls, and achievement fantasy correlated with IQ gain, it was suspected that early maternal concern with achievement would be positively correlated with IQ gain in girls but not necessarily with boys. These assumptions were verified, for the product-moment correlation for girls was $+.37$ ($p < .05$), while for boys it was $-.10$ ($p > .20$).

DISCUSSION

The results indicate that achievement fantasy shows more than chance stability during childhood and early adolescence. The data, therefore, add validity to predictive and explanatory statements about achievement behavior derived from fantasy measures.

With respect to antecedent factors, it seems

that for this select middle-class sample achievement fantasy was not significantly related to the general level of education in the family but partly dependent on the achievement concern of the same-sex parent, for maternal concern with achievement correlated only with achievement themes in girls. It is suggested that concern with achievement behavior and the sex of the parent showing the concern may interact in influencing the acquisition of achievement-strivings in the child.

At a more speculative level, it is possible that an opposite-sex parent's concern with achievement may lead to some inhibition of achievement-striving in the child if he perceives achievement behavior as an attribute of the opposite sex. Such speculations agree with psychoanalytic hypotheses to the effect that learning problems often arise when the child associates intellectual achievement with the goals and attributes of the opposite-sex parent (Blanchard, 1946; Pearson, 1952), and the consequent anxiety over sex role identity inhibits learning and intellectual effort.

There are several cautions with respect to generalizing from these data. First, the specific fantasy stimuli used tended to elicit themes emphasizing success and the acquisition of achievement goals, rather than failure or anxiety over not attaining them. Clark, Teevan, and Ricciuti (1956) have reported that the relationships between achievement fantasy and a criterion of overt achievement were quite different when the fantasy stimuli suggested success stories from when they suggested failure. Second, the stimulus that accounted for most of the achievement themes was Card 1 (boy with violin), which typically evoked the theme of intellectual mastery of the instrument. It is possible that if there were no fantasy stimulus suggesting achievement on an intellectual task the relationship with IQ gain and level of parental education might not have appeared. It is suggested that the greater the comparability in content between the fantasy and overt behavior, the greater the predictive power of fantasy.

The occurrence of any achievement-related behavior, be it IQ gain or a TAT theme, it is suggested, is a result of the interaction of (a) the intensity of the motivation for a specific achievement goal, (b) the strength of response

tendencies directed toward attaining it, and (c) the strength of response tendencies that oppose striving toward attaining the goal. Motive is defined, after McClelland, as the anticipation of a specific goal. However, this anticipatory state does not imply any specific cognitive or instrumental response associated with it, and the specific behaviors that follow arousal of a motive depend, in part, on the strength of goal-directed responses and the strength of anxiety-based avoidant responses associated with the motive. Two potential sources of anxiety that may be associated with achievement goals are fear of failure and association of intellectual achievement with opposite-sex attributes. The strength of these anxiety-based avoidant responses may interact with the approach-response tendencies in influencing the occurrence of both direct achievement behavior and achievement fantasy. Achievement fantasy is, therefore, viewed as one index of the strength of the response tendency to strive for a specific achievement goal.

Recent research (Kagan, 1956; Mussen & Scodel, 1955) has seriously questioned the validity of the prevalent assumption that a projective stimulus minimizes the effects of anxiety on the expression of a motive and thus provides an index of "pure" motive strength. Since the stimuli used to obtain achievement themes usually suggest achievement content it seems plausible that if an S has strong anxiety over achievement goals, achievement themes may not appear in the protocol despite strong achievement motivation. At present there are no well-validated techniques that allow an investigator to use various aspects of the person's contemporaneous behavior to assess motivation as well as approach and avoidant response tendencies. One solution to this problem may lie in the use of antecedent, historical data as indices of motivation.

SUMMARY

The following data on 44 boys and 42 girls in the Fels research population were studied to assess the reliability and validity correlates of achievement fantasy.

a. Three TAT protocols administered at ages 8-9, 11-6, and 14-6.

b. Educational level of the parents.

c. Ratings of maternal concern with achievement for the first three years of the child's life.

d. Changes in IQ score during the years 6 through 10.

The results showed (a) an increase in occurrence of achievement themes with age and better than chance stability of occurrence over time, (b) a positive relation between early maternal concern with achievement and both achievement fantasy and IQ gain in girls, (c) a nonlinear but generally positive relation between occurrence of achievement fantasy and IQ gain for boys and girls, and (d) a suggestive relation between achievement fantasy and educational level of the same-sex parent.

It was suggested that achievement fantasy is an index of the strength of the subject's behavioral tendency to seek achievement goals and that achievement concern on the part of the same-sex parent has a special influence on the child's achievement strivings. Specific methodological and theoretical problems with respect to measurement and interpretation of achievement fantasy were discussed.

REFERENCES

- ATKINSON, J. W. Explorations using imaginative thought to assess the strength of human motives. In M. R. Jones, *Current theory and research in motivation: A symposium*. Lincoln: Univer. Nebraska Press, 1954. Pp. 56-112.
- ATKINSON, J. W. Motivational determinants of risk-taking behavior. *Psychol. Rev.*, 1957, **64**, 359-372.
- BLANCHARD, PHYLLIS. Psychoanalytic contributions to the problems of reading disabilities. In *Psychoanalytic study of the child*. II. New York: Int. Univer. Press, 1946. Pp. 163-188.
- CLARK, R. A., TEEVAN, R., & RICCIUTI, H. N. Hope-of-success and fear-of-failure as aspects of need for achievement. *J. abnorm. soc. Psychol.*, 1956, **53**, 182-186.
- FARBER, I. E. Comments on Professor Atkinson's Paper. In M. R. Jones, *Current theory and research in motivation: A symposium*. Lincoln: Univer. Nebraska Press, 1954. Pp. 112-115.
- FRENCH, ELIZABETH G. Some characteristics of achievement motivation. *J. exp. Psychol.*, 1955, **50**, 232-236.
- FRENCH, ELIZABETH G. The interaction of achievement motivation and ability in problem-solving success. *J. abnorm. soc. Psychol.*, 1958, **57**, 306-309.
- FRENCH, ELIZABETH G., & THOMAS, F. H. The relation of achievement motivation to problem-solving

- effectiveness. *J. abnorm. soc. Psychol.*, 1958, **56**, 45-48.
- KAGAN, J. Measurement of overt aggression from fantasy. *J. abnorm. soc. Psychol.*, 1956, **52**, 390-393.
- KAGAN, J., SONTAG, L. W., NELSON, VIRGINIA L., & BAKER, C. T. Personality and IQ change. *J. abnorm. soc. Psychol.*, 1958, **56**, 261-266.
- LAZARUS, R. S., & BAKER, R. W. Motivation and personality in psychological stress. *Psychol. Newsltr.*, N.Y.U., 1957, **8**, 159-193.
- LOWELL, E. L. A methodological study of projectively measured achievement motivation. Unpublished master's thesis. Wesleyan Univer., 1950.
- MCCLELLAND, D. C., & LIBERMAN, A. M. The effect of need for achievement on recognition of need-related words. *J. Pers.*, 1949, **18**, 236-251.
- MCCLELLAND, D. C., ATKINSON, J. W., CLARK, R. A., & LOWELL, E. L. *The achievement motive*. New York: Appleton-Century-Crofts, 1953.
- MUSSEN, P. H., & SCODEL, A. The effects of sexual stimulation under varying conditions on TAT sexual responsiveness. *J. consult. Psychol.*, 1955, **19**, 90.
- PEARSON, G. H. J. A survey of learning difficulties in children. In *Psychoanalytic study of the child*. VII. New York: Int. Univer. Press, 1952. Pp. 322-356.
- SONTAG, L. W., BAKER, C. T., & NELSON, VIRGINIA L. Mental growth and personality development. *Monogr. Soc. Res. Child Developm.*, 1958, **23**, No. 68.
- VOGEL, W., LAZARUS, R. S., & BAKER, R. W. A comparison of fantasy and behavioral indices as predictors of extrinsically motivated behavior under conditions including a failure stress. *Amer. Psychol.*, 1957, **12**, 400.
- WINTERBOTTOM, M. R. The relation of childhood training in independence to achievement motivation. Unpublished doctoral dissertation, Univer. of Michigan, 1952.

Received February 7, 1958.

NEED AFFILIATION AND FREQUENCY OF FOUR TYPES OF COMMUNICATION

JOHN B. LANSING AND ROGER W. HEYNS¹

University of Michigan

THIS paper reports an effort to predict social behavior from a personality variable. Specifically, it is another study in a series aimed at discovering the behavioral correlates of need for affiliation. Shipley and Veroff (1952) developed a projective measure of need for affiliation, defining the motive primarily in terms of separation anxiety and concern over rejection. Atkinson, Heyns, and Veroff (1954) in a similarly designed study constructed a measure based upon a more positive definition: *concern with the establishment, maintenance, or restoration of a code of personal relationship, best symbolized by the word friendship*. French and Chadwick (1956) have recently reported results that are consistent with this more positive conception of affiliation motivation. It is this definition which is used in the present study.

Data concerning the empirical correlates of the need for affiliation are still sparse, since the methods of measurement are recent. French (1956), using a somewhat different measure but the same conception of need for affiliation, has reported results which indicate that the subject's choices of a work partner are correlated with need for affiliation and need for achievement. Subjects with high need for affiliation and low achievement motivation pick a less competent friend for a work partner, while subjects with high need for achievement and low affiliation motivation choose a competent nonfriend. In the validation study, Atkinson et al. (1954) report that the persons with the highest need for affiliation were more likely to be cited by their friends as seeking approval. Both the Shipley-Veroff (1952) and the Atkinson et al. (1954) studies report a negative correlation between need for affiliation and popularity in the group. The present study is an attempt to extend our knowledge of the

relationship between the need for affiliation and other social acts.

The data reported here were collected in an investigation of motives, attitudes, and long distance telephone calls. In the study, a number of independent variables were measured which might be related to use of the long distance telephone for social calls; several alternative channels of communication were also investigated. The present report is concerned with the relationship between n Affiliation and frequency of use of these four forms of communication: (a) use of long distance telephone for social calls, (b) use of the local telephone, (c) writing letters, and (d) visiting close friends and relatives living at a distance.

The initial hypothesis held that people who score high on n Affiliation also tend to be frequent users of each of these methods of communication, other things being equal. Among the relevant other factors to be controlled (discussed below) is, for example, income: a person's n Affiliation should be expected to predict the frequency of his social long distance calling only when his income is taken into account.

This research is unusual in that it attempts to quantify the relation between a measure of personality and the behavior of a segment of the general population. Published studies known to us for the most part are concerned with more restricted populations, commonly populations of college students. Also, the present study differs from laboratory investigations in that the behavior measured is behavior in "real life," uninfluenced by the investigation and measured after the event. The methods used here have their difficulties and are not necessarily superior. They are, however, of interest as a complement to other techniques. One objective of the study of personality, in our view, is to measure dimensions of personality that can be related quantitatively to the behavior of the population at large. We all tend to apply research results to

¹ The authors wish to acknowledge their indebtedness to their colleagues at the Survey Research Center, particularly to Charles Cannell, Joseph Adelson, and John Colombotos who were closely associated with the larger research project that is reported here in part.

the general population in our own thinking. It seems to us that there is a need for research that attempts systematically to carry through this process of generalization.

METHOD

The sample used in this research was designed especially for the study of frequency of use of residential telephones for long distance calls. The universe sampled consisted of residential subscribers of the Michigan Bell Telephone Company and their wives, in the southern part of the lower peninsula of the state of Michigan. A sample of 1571 subscribers was drawn, and information was tabulated from company records about the frequency of use of long distance by each. For purposes of this study, a "long distance" call was defined as a call to a point over 50 miles from the point of origin. At this point three groups were defined, "low" users of long distance, "intermediate" users, and "high" users. A "high" user is defined to be a subscriber who made six or more long distance calls in the last period of three months for which data were available at the time when the sample was selected (in the early fall of 1955). A "low" user was a subscriber who made no long distance calls in the same period. Personal interviews were taken with "low" users and "high" users but not with "intermediate" users. Although the number of low users in the population is much larger than the number of high users, half the interviews were taken with low users and half with high users, to permit efficient comparison of the groups.

Once a subscriber was selected for interview, if he was married his wife also was designated to be interviewed. If the subscriber had no spouse, only the subscriber fell in the sample. Interviews were taken with 400 individuals representing 230 families.

This sample design, using as it does extremes in the behavior of immediate interest, is efficient for the investigation of factors related to the frequency of long distance calling. It is of uncertain efficiency for the investigation of factors relating to the frequency of other forms of communication. Estimates of relationships between independent variables and these other types of communication will be in error if the independent variables enter into different relationships among people who are intermediate in their use of long distance from those that are characteristic of high and low users. We have no reason to believe that the effect of the independent variables is complicated in this manner.

The sampling procedure does not permit estimation of the frequency of different types of communication among the population as a whole. For example, the data should not be used to estimate the mean number of letters written per year per adult in Michigan.

The interviewing techniques were similar for the most part to those used in other cross-section studies by the Survey Research Center (Campbell & Katona, 1953). The interviews lasted about an hour. A common pattern was for the interviewer to interview the wife during the day and return in the early evening to interview the husband.

The Affiliation motive was measured by means of responses to TAT pictures. Each respondent was shown a series of four pictures and asked to tell a story about each.² The general procedure followed here has been described in detail elsewhere (Atkinson et al., 1954). Some of the pictures that had been satisfactory with college students were not successful, largely because the persons in the pictures were too youthful.³ The interviewers received special instruction in the use of the pictures, and had no serious difficulty in using them in the interview. Training was directed particularly to overcome a tendency observed in the pretest for protocols to be too brief to be scored properly. The scoring of the protocols was done by the second author, according to the published manual (Heyns, Atkinson, & Veroff, 1958). Before scoring the present material he checked his reliability by scoring 50 stories that had been set aside for purposes of training and reliability assessment. The total scores he assigned to the 50 stories correlated .94 with those which had previously been ascribed to them.

The techniques used in the analysis of the data include the familiar frequency tabulations and multiple regression. The estimates of sampling error for the regression coefficients shown were calculated by a technique that assumes that the sample used was a simple random sample. In fact, the sample is clustered. That is, the units in the sampling were not selected independently.

The problems of estimating sampling errors of clustered samples have been discussed elsewhere (Kish, 1957). The essential difficulty is that the usual formulas lead to underestimates of the true sampling errors when applied to clustered samples such as the present one. The drawing of the present sample proceeded in stages. In the interests of economy, interviews were taken only in certain telephone exchanges in southern Michigan instead of being spread all over that part of the state. Our sample was large enough so that we took interviews in all metropolitan areas, and these covered about two-thirds of our population. As a result, though we do not have quite as much information as we would if the interviews had been spread more widely, the usual formulas for estimating the sampling

² The sequence of questions used was as follows:

- I'm going to show you some pictures and I'd like you to make up a story about them.
- What's going on here?
 - What is the person (what are they) thinking?
 - What do they want?
 - What's going to happen next?

³ The pictures finally chosen were as follows: For women (a) a young girl standing in a transportation terminal, (b) three young women talking over tea cups, (c) two men and two women around a desk in an office, and (d) huddled figure of young person with head on a cot. For men (a) two men, one sitting on a fence, rural background, (b) a male passenger on an airplane, (c) four men sitting and lounging informally in a furnished room, and (d) two men standing, facing each other, talking.

TABLE 1

RELATIONSHIP BETWEEN NEED AFFILIATION SCORE
AND NUMBER OF LOCAL CALLS MADE
OR RECEIVED "LAST" WEEK^a

Number of Calls	All Respondents	Need Affiliation Score		
		Low (00-03)	Middle (04-06)	High (07-19)
Up to 2	18	21	21	10
3 to 14	49	52	51	45
15 and over	29	22	26	41
Not ascertained	4	5	2	4
Total	100	100	100	100
Number of respondents	400	155	103	125

^a The question was: "About how many calls in the last week have you yourself made or received on your home phone?" The question is phrased to include all types of calls, but answers may be assumed to refer primarily to local calls.

error of proportions can be used with minor adjustments.⁴

A more serious complication arises because both husband and wife were interviewed in the family. If husband and wife behave as a unit with respect to the behavior under study, we do not have as many independent observations as we have interviews. In fact, we have only as many independent observations as we have married couples in the sample. If husband and wife behave independently, however, the number of independent observations is the same as the number of interviews. Thus, where there is no correlation between data from the interview with the husband and data from the interview with the wife, the number of interviews is equivalent to the number of observations and should be used in entering the table above. When perfect positive correlation exists, the number of observations is equal to the number of married couples.

Regression analysis requires that the variables used be scaled. As will be apparent from a study of the variables used, some of the scales imposed on the data were defined by the use of judgment. To the extent that the judgment used was poor, the effect is to reduce the chances of obtaining significant results. Operationally, our hypotheses are to the effect that the independent variables as scaled will have an effect on the dependent variables as scaled.

⁴ Differences equal to or greater than those shown below may be taken as significant at the 95% level:

Approximate Value of the Percentage Found	Minimum Difference Between Two Groups Needed for Significance	
	For two groups each of 200	For two groups each of 100
35 to 65%	11	15
20 or 80%	9	12
10 or 90%	7	—

TABLE 2

RELATIONSHIP BETWEEN NEED AFFILIATION SCORE
AND NUMBER OF TIMES A YEAR RESPONDENT
GETS TOGETHER WITH OUT-OF-TOWN
FRIENDS OR RELATIVES^a

Number of Visits	All Respondents	Need Affiliation Score		
		Low (00-03)	Middle (04-06)	High (07-19)
Up to 1	45	51	49	36
1 to 10	36	31	36	39
11 or more	10	9	6	14
Other answers or not ascertained	9	9	9	11
Total	100	100	100	100
Number of respondents ^b	308	118	80	101

^a The question was: "Thinking of your half dozen closest friends or relatives, do they all live here in (name of city), or do they live in other places, or what? About how often do you get together with these people?"

^b Number of respondents includes only those respondents who have relatives or friends living 50 miles or more away.

RESULTS

Frequency Analysis

Tables 1 to 4 present the basic data concerning the relationship of n Affiliation to frequency of each of four kinds of communication in turn.

The relation between n Affiliation score and number of local calls made or received by the person "last" week (Table 1) is statistically significant and in the predicted direction. There is no necessary correlation between the number of calls on the home phone by husband and wife, but we may allow for clustering by conservatively estimating that we have 100 low and 100 high scores on n Affiliation instead of the 155 and 125 observed. We require for statistical significance a difference of 14 percentage points between the proportion of low scorers making 15 or more calls and the proportion of high scorers making 15 or more calls. The obtained difference is 19 percentage points.

In the next three tables, which concern communication with friends and relatives living 50 miles or more away, we omit respondents who report that none of their half dozen closest friends and relatives live at that distance.

Table 2 shows a relationship between n

TABLE 3
RELATIONSHIP BETWEEN NEED AFFILIATION SCORE
AND NUMBER OF LETTERS WRITTEN PER
YEAR TO RELATIVES OR FRIENDS^a
(Percentage distribution of respondents)

Number of Letters	All Re- spond- ents	Need Affiliation Score		
		Low (00-03)	Middle (04-06)	High (07-19)
Up to 20	56	56	55	52
21 to 75	18	20	16	18
76 or more	21	15	25	26
Other answers or not as- certained	5	9	4	4
Total	100	100	100	100
Number of respond- ents ^b	308	118	80	101

^a The question was: "Thinking of your half dozen closest friends and relatives, do they all live here in (name of city), or do they live in other places, or what? About how many letters would you say you write to relatives or friends in a year?"

^b Number of respondents includes only those respondents who have relatives or friends living 50 miles away or more.

Affiliation score and number of visits that is in the predicted direction but too small for significance if we allow for clustering.

Table 3 shows a relationship between n Affiliation score and number of letters written per year that is also below the margin of statistical significance but in the predicted direction. Table 4 shows no relation between n Affiliation score and frequency of use of long distance which is probably the most complexly determined of the four types of communicative acts. The interpretation of this result is discussed in some detail later.

Thus, our two-way tabulations show a complicated picture. There is a clearly significant relation between n Affiliation and the use of the local phone. The relationship of the need to frequency of letters and visits is in the predicted direction but is not significant. Finally, there is no relationship between the need measure and use of the long distance phone. We turn to more powerful analytical methods.

Regression Analysis

Tables 5 and 6 define the variables used in regression analysis of visits and frequency of letter writing, respectively. The independent variables in the two equations are identical. Family income (X_1) was introduced on the

TABLE 4
RELATIONSHIP BETWEEN NEED AFFILIATION SCORE
AND FREQUENCY OF USE OF
LONG DISTANCE
(Percentage distribution of respondents)

Frequency of Calling	All Re- spond- ents	Need Affiliation Score		
		Low (00-03)	Middle (04-06)	High (07-19)
Low Users	43	37	50	46
High Users	57	63	50	54
Total	100	100	100	100
Number of re- spondents ^a	308	118	80	101

^a Number of respondents includes only those respondents who have relatives or friends living 50 miles or more away.

hypothesis that high income people are more likely to make visits at a distance and to write letters. X_2 is a measure of the closeness (degree of kinship) of the relatives living away. The hypothesis is that those with only distant relatives away are less frequent visitors and less prolific letter writers than those whose close relatives live away. Distance (X_3) forms a barrier to visits, and, we predicted, may indirectly weaken all ties, reducing even letter writing. Education (X_4) as a measure of socioeconomic status may have an effect on visiting, on the hypothesis that people with high education travel more, even for a given income level. Education, as a measure of familiarity with the written word, should be related to frequency of letter writing. Occupation (X_5) was introduced as an additional measure of socioeconomic status, with the direction of prediction similar to that for education. Need Affiliation scores ranging from 0 to 19 enter as X_6 .

The results for visits are as follows:

$$[1] \quad Y_1 = 3.44 + .05X_1 - .02X_2 - .29X_3 + .01X_4 + .11X_5 + .049X_6$$

$R = .363$ $R^2 = .13$

(.04)

(.06)

(.05)

(.06)

(.06)

(.025)

Under each coefficient, in parenthesis is its standard error (one sigma). Our interest is primarily in n Affiliation (X_6), which shows a coefficient equal to two times its standard error and in the predicted direction. As noted above, the sample design is such that the

TABLE 5

DEFINITION OF VARIABLES FOR REGRESSION ANALYSIS OF FREQUENCY OF VISITS, EQUATION [1]

Variable	Symbol	Range of Values
Frequency of visits	Y_1	1. Less than once a year 2. Once a year; "very irregularly" 3. 1-5 times a year; frequency of visits not ascertained 4. 6-10 times a year 5. 11-15 times a year; "fairly frequently," "regularly" 6. 16 or more times a year
Family income	X_1	1. Under \$1000 2. \$1000-1999 3. \$2000-2999 4. \$3000-3999 5. \$4000-4999; income not ascertained 6. \$5000-5999 7. \$6000-7499 8. \$7500-9999 9. \$10,000 or more
Closeness of relatives living away	X_2	0. Has no relatives living away (Interviews coded 0 on this scale were not included in the regression) 1. Has only distant relatives away 2. Has both close and distant relatives away 3. Has only close relatives away
Distance friends and relatives live away	X_3	1. 50-299 miles (midpoint 175 miles) 2. 50-299 and 300-600 (midpoint 312 miles) 3. 300-600 only; 50-299 and over 600; not ascertained (midpoint 470 miles) 4. 300-600 and over 600 (midpoint 625 miles) 5. Over 600 miles (midpoint 800 miles)
Education of head of family	X_4	1. Attended grammar school (1-8 yrs.) 2. Attended high school (9-11 yrs.); 1-8 yrs. plus other education 3. Completed high school; education not ascertained 4. Attended college, no degree 5. College degree
Occupation of head of family	X_5	1. Laborer, service worker 2. Craftsman, foreman, operative, member of armed forces 3. Housewife, retired; occupation not ascertained 4. Clerical, sales 5. Professional, managerial, self-employed
Need affiliation	X_6	00. - 19. (actual score)

TABLE 6

DEFINITION OF DEPENDENT VARIABLE FOR REGRESSION ANALYSIS OF FREQUENCY OF LETTER WRITING,^a EQUATION [2]

Variable	Symbol	Range of Values
Frequency of letter writing	Y_2	0. None at all 1. 1-5 letters; "just a few" 2. 6-20 letters 3. 21-40 letters, number not ascertained 4. 41-50 letters 5. 51-75 letters 6. 76 letters or more

^a The independent variables in Equation [2] are identical with those in Equation [1]. See Table 5.

standard errors shown are underestimates of the true errors. If we were dealing with a simple random sample, the probability of such an event occurring by chance would be about .03. We may look at the matter in another way by saying that a departure of 1.65 sigmas from zero in a predicted direction will occur by chance 5% of the time. Our departure from zero is .049. If the true sigma is about .03 or smaller, our result is significant at the 5% level. We suspect that the effect of clustering is powerful enough to raise the standard error from .025 to more than .03. We can say only that the data seem to point in the direction indicated by the hypothesis.

In interpreting the results for the other variables we may consider jointly X_1 (income), X_4 (education), and X_5 (occupation). These variables are intercorrelated, with simple correlation coefficients ranging from .37 (X_1 with X_4) to .58 (X_4 with X_5). In view of these two-way coefficients, it is not surprising that no one of the variables has a significant incremental effect. The distance people live away (X_3) is shown to have an effect on the frequency of visits. Degree of kinship of relatives (X_4) has no effect.

The results for letters are as follows:

$$[2] \quad Y_2 = 2.19 + .01X_1 - .19X_2 - .10X_3 \\ \quad \quad \quad (.07) \quad (.11) \quad (.09) \\ \quad \quad \quad + .10X_4 + .10X_5 + .84X_6 \\ \quad \quad \quad (.11) \quad (.11) \quad (.36)$$

$$R = .126$$

$$R^2 = .02$$

Again, we are interested primarily in *n* Affiliation (X_6). Again, the results would be significant without question if we were dealing with a simple random sample, since the coefficient of .84 is more than twice its standard error (.36) and is, furthermore, in the direction predicted by the theory. We have no reason to believe that husbands and wives write letters jointly as commonly as they visit distant relatives together. In fact, we would anticipate that as often as not if one of the two is an enthusiastic letter writer the other may write fewer letters. Hence, it seems to us unlikely that the true standard error is enough larger than .36 to change our conclusions. (An error of .51 or above would be needed.) Thus, as best we can judge, the data support the hypothesis. We note, however, that the multiple correlation coefficient is only .126, and its square, only .02.

Of the other independent variables considered, none show any effect. Only X_2 (closeness of relative living away) approaches significance at all, and here the sign of the coefficient is the reverse of that predicted.

Finally, a multiple regression was computed, not presented here, using as dependent variable whether the individual (or spouse) was a low user or a high user of long distance and using score on *n* Affiliation as one of the independent variables. The results as far as *n* Affiliation is concerned were as unpromising as those in Table 4.

We did not compute a multiple regression for local telephone calls. As a by-product of the calculations described above we computed two-way correlation coefficients between each pair of independent variables, and found none of our demographic variables to be correlated with *n* Affiliation. Thus, there was no possibility that the correlation observed in Table 1 could be attributed to an association between *n* Affiliation and one of the demographic variables even if one of the latter proved to be related to the number of local calls.

DISCUSSION AND FURTHER ANALYSIS

The analysis up to this stage left the investigators with an unresolved problem. Need Affiliation as measured predicts local calling and letter writing and may be related even to

visiting people at a distance. Why, then, does it not predict social long distance calling?

One possible answer is that *n* Affiliation is not relevant at all to long distance telephoning. We have tested this hypothesis by examining the relation between *n* Affiliation and other scores that bear on long distance calling. One of these was a measure of respondents' feelings while making long distance calls (Table 7). A second score was obtained from answers to a direct question as to whether people enjoy making long distance calls (Table 8).

The data show little relation between *n* Affiliation and the score on feelings while making a call. (This score, it should be added, is related to frequency of use of long distance.) The data do, however, show a relation between his score on *n* Affiliation and whether a respondent says he enjoys making long distance calls. From these data we conclude that *n* Affiliation as measured is not totally irrelevant to use of the long distance phone. At a minimum it does have something to do with people's reports of whether they enjoy making long distance calls.

There are several possible explanations of why *n* Affiliation might be relevant and yet show no direct relationship to frequency of long distance in our data. The first hypothesis is the

TABLE 7
RELATIONSHIP BETWEEN NEED AFFILIATION SCORE
AND SCORE ON RESPONDENT'S FEELINGS
WHILE PLACING A LONG DISTANCE
CALL^a

Feelings	All Respondents	Need Affiliation Score		
		Low (00-03)	Middle (04-06)	High (07-19)
Pleasant feelings	59	58	54	69
About half and half—neutral	23	25	26	15
Unpleasant feelings	11	13	13	9
Not ascertained	7	4	7	7
Total	100	100	100	100
Number of respondents ^b	347	132	87	115

^a The question was: "Here is a list of ways people tell us they feel when they make a long distance call. Do any of these words describe how you sometimes feel when making a long distance call? Relaxed? Nervous? Tense? Anxious? Happy?" Scoring involved assigning +1 to each answer indicating presence of a positive feeling or absence of a negative feeling.

^b Does not include those respondents who have never made a long distance call.

TABLE 8
RELATIONSHIP BETWEEN NEED AFFILIATION SCORE
AND FEELINGS ABOUT MAKING
LONG DISTANCE CALLS^a
(Percentage distribution of respondents)

Feelings	All Respondents	Need Affiliation Score		
		Low (00-03)	Middle (04-06)	High (07-19)
Enjoy it; enjoy it moderately	60	55	61	70
Sometime enjoy, sometime dislike; neither enjoy nor dislike	27	33	25	22
Do not enjoy	6	7	6	1
Not ascertained	7	5	8	7
Total	100	100	100	100
Number of respondents ^b	347	132	87	115

^a The question was: "How do you feel about making a long distance call—do you enjoy it or dislike it or what?"

^b Does not include those respondents who have never made a long distance call.

simplest: that the effect of *n* Affiliation is too small to detect, using our data. In support of this contention one can call attention to the small size of the sample and the difficult problems of measurement. Weaknesses in the *n* Affiliation measure would be even more serious if the affiliative motive has only a slight impact. One may note further that the available data from the telephone company refer to the total number of calls placed per telephone, while the *n* Affiliation scores are for individuals. A variation on this argument is to take the position that only those who score very high on *n* Affiliation are different from the rest of the population. According to this view, most of the observed variation in affiliation score is irrelevant. And one may recall that the observed effects of *n* Affiliation on visiting and letter writing are at best small. All of these points seem to us to add up to a plausible case that there is an effect, but it is too small for us to detect.

A second hypothesis is more complex. It is that the effect of *n* Affiliation depends in some complex way upon the effect of other variables. The need for affiliation, according to this view, has its predicted effect only in certain circumstances; indeed, its effect may even be the reverse from one set of circumstances to another. Or, it may be associated with other

variables whose effect varies from one set of circumstances to another.

Another possible area of speculation is that of operation of the affiliative motive itself. There is a clear relationship of the motive measure to communication acts such as local calls, letter writing, and a very suggestive one to visits. These acts may be successful enough in the lives of our respondents to satisfy the affiliative need, leaving only distance calling to be influenced by other factors. One would expect a relationship between the motive measure and long distance calls if these other acts were unsuccessful.

A fourth possibility concerns the measure itself. The definition of affiliative imagery and the scoring instructions are oriented toward friendship as the prototype of the affiliative relationship. Statements of family ties are not automatically scored as affiliative; there must be evidence of concern about the relationship. This is a stringent requirement; it means that the measure is less slanted toward ties between relatives. Our data show, however, that social long distance calls are primarily to relatives. It may be that a measure more sensitive to the desire for close ties to relatives would show a greater relationship. The same reasoning would argue that the present measure might predict amount of calling to friends. Unfortunately, it was not possible to explore this possibility.

This discussion has emphasized the failure of the Affiliation measure to predict social long distance calling because of the intriguing theoretical and methodological issue raised by the missing relationship. It is important to note, however, that this kind of speculation is a reasonable activity only because of our most important finding, that the *n* Affiliation measure is significantly related to some other communication acts, particularly local calling and letter writing and perhaps to making visits to friends.

SUMMARY

This article reports the results of an attempt to relate need for affiliation to the frequency of four types of communication: (a) use of long distance telephone for social calls, (b) use of the local telephone, (c) writing letters, and (d) visiting close friends and relatives living at a

distance. The data were collected in 400 personal interviews with a specially designed sample of telephone subscribers living in southern Michigan. The research is unusual in that it concerns the relation between a dimension of personality, measured by TAT-type pictures, and the behavior of a segment of the general population.

The major findings are that n Affiliation is correlated in the predicted direction with use of the local telephone. The same relationship is found to letter writing, though the correlation is low. With regard to visits, the correlation is in the predicted direction but not significant. Need affiliation as measured does not predict social long distance calling.

REFERENCES

ATKINSON, J. W., HEYNS, R. W., & VEROFF, J. The effect of experimental arousal of the affiliation

motive on thematic apperception. *J. abnorm. soc. Psychol.*, 1954, **49**, 405-410.

CAMPBELL, A. A., & KATONA, G. The sample survey. In L. Festinger & D. Katz (Eds.), *Research methods in the behavioral sciences*. New York: Dryden, 1953.

FRENCH, ELIZABETH G. Motivation as a variable in work partner selection. *J. abnorm. soc. Psychol.*, 1956, **53**, 96-99.

FRENCH, ELIZABETH G., & CHADWICK, IRENE. Some characteristics of affiliation motivation. *J. abnorm. soc. Psychol.*, 1956, **52**, 296-300.

HEYNS, R. W., ATKINSON, J. W., & VEROFF, J. A scoring manual for the affiliation motive. In J. W. Atkinson (Ed.), *Motives in fantasy, action and society*. Princeton: Van Nostrand, 1958.

KISH, L. Confidence intervals for clustered samples. *Am. sociol. Rev.*, 1957, **22**, 154-165.

SHIPLEY, T. E., & VEROFF, J. A projective measure of need for affiliation. *J. exp. Psychol.*, 1952, **43**, 349-356.

Received February 12, 1958.

RE-EVALUATION OF CHOICE ALTERNATIVES AS A FUNCTION OF THEIR NUMBER AND QUALITATIVE SIMILARITY¹

JACK W. BREHM² AND ARTHUR R. COHEN

Yale University

MOST theory and research relevant to choice situations have centered on prechoice phenomena. There is, however, increasing concern with postdecision phenomena. A theory recently proposed by Festinger (1957), for example, allows several specific derivations about the psychological consequences of choice. Some of these derivations have been tested (Brehm, 1956; Ehrlich, Guttman, Schonbach, & Mills, 1957; Festinger, 1957). This paper reports a test of two further derivations.

The central conception of this theory is that a person who holds cognitions about himself or the environment that are inconsistent with each other experiences psychological "dissonance," which he attempts to reduce or eliminate by trying to reduce the number of inconsistent cognitions in relation to the number of consistent ones. An approximate statement of the fundamental hypothesis as it applies to the choice situation is that *the magnitude of psychological dissonance is a direct function of what one has to give up compared to what one obtains*. That is, the attractiveness of an unchosen alternative is a cognition inconsistent with the knowledge that one has chosen a different alternative. In purchasing a new television receiver, for example, having to give up a new car would create more dissonance than having to give up a new pair of shoes.

This general statement allows two further derivations about the consequences of choice: (a) the greater the qualitative dissimilarity of alternatives, the greater the magnitude of dissonance from choice; and (b) the greater the number of alternatives, the greater the magnitude of dissonance from choice. Other things being equal, as qualitative dissimilarity increases, what one has to give up relative to what one gains increases. For example,

choosing one of two automobiles of different makes (but of equal attractiveness) would create more dissonance than would choosing one of two that are identical except for color. More features such as type of styling, type of motor, etc., must be given up by a choice between different makes than by a choice between two nearly identical cars of the same make. The number of alternatives affects the magnitude of dissonance in much the same way. Where the alternatives have at least some qualitative dissimilarity, and with other factors held constant, the greater the number of alternatives from which one must choose, the more one must give up and consequently the greater the magnitude of dissonance.

When a person experiences psychological dissonance, he tries to reduce or eliminate it. If the dissonance has occurred as the consequence of making a choice, it may be reduced or eliminated by re-evaluating the alternatives so as to minimize what has been given up *relative* to what has been obtained. Thus to the extent to which an individual is experiencing dissonance, he may be expected to increase the perceived attractiveness of the chosen alternative and decrease the perceived attractiveness of unchosen alternatives.

The experiment described was designed to test the following two general hypotheses:

1. Other things being equal, the greater the qualitative dissimilarity of alternatives, the greater is the magnitude of dissonance created by choice, and the greater the consequent re-evaluation of these alternatives.

2. Other things being equal, the greater the number of alternatives, the greater is the magnitude of dissonance and the consequent re-evaluation of these alternatives.

METHOD

The strategy was to ask sixth-grade children to indicate, both before and after a choice, how much they liked each of several different toys. The choice given to about one half of the Ss was between similar toys, and to the other half it was between dissimilar toys. From each of these two groups about half of the Ss were given a choice between two toys, while the

¹ This is a report of a study conducted as part of the Yale Communications Research Program, and supported by grants to Carl I. Hovland from Bell Telephone Laboratories and from the Rockefeller Foundation.

² Now at Duke University.

other half were given a choice of one among four toys. Ss' average change in liking of the chosen and unchosen alternatives was then computed for each kind of choice.

Subjects. The original sample consisted of 96 boys and 107 girls in eight different classes in three different schools,³ who were present for both the first and second sessions of the experiment. A selection procedure, to be described later, reduced the usable number to 29 boys and 43 girls.

Procedure. The experimenter (*E*) and his helper were introduced to each class as persons from Yale University who were doing a survey having to do with toys. *E* then said to the Ss:

Some toy manufacturers have recently asked us to find out what kinds of toys people like you like and what kinds you don't like. Christmas is not very far away and so it's especially important for the toy manufacturers to know what you like or dislike so they will know what to make and send to all the stores.

Now in order to find what you really do like or dislike we are going to ask each of you to indicate on a questionnaire just how much you like or dislike each of 16 different toys.

After the questionnaires had been distributed and explained, the 16 toys, all in one large box, were brought into the room. Each toy had a number to correspond with the number on one of the rating scales on the questionnaire. *E* then went through the toys by numbered order, holding each long enough and close enough so that it could be inspected closely by every child. At the beginning of this procedure Ss were told that their own personal answers were wanted, not what their neighbor thought, so they were asked not to talk while filling out the questionnaire. One or two gentle reminders were usually necessary during the course of the rating. When the ratings were completed, the questionnaires were collected and the Ss thanked for helping with the survey. No mention was made of the possibility that *E* and his helper would return.

The second session took place exactly one week after the first. It was conducted by the same *E*, though for some classes a different person had to be used as the helper.⁴ *E* first reminded the Ss that he and his helper were from Yale and were conducting a survey for some toy manufacturers. He then went on to say:

The information you gave us last week was very good. In fact, it was so good that the toy manufacturers have asked us to get some more information. So today we're going to ask you to help us out by answering some more questions about the toys.

Now since the toy manufacturers were so pleased

with what you did for them last week, and since they are asking you to do something more for them this week, they thought it only fair that they give you something for all this help. So they decided to give each one of you one of the toys that you've been rating. Unfortunately, however, we don't have enough of any one toy so that you could all have the same thing. That means we can't let you choose just anything you want. So what we did was to pick two (or four) for each of you to choose between. While we've tried to give you a choice between things you like, you probably won't get a chance to choose the one you like most. We just couldn't arrange the choices that well.

Ss were then given prearranged choice sheets that indicated the toys, by name and number, from which they could choose. They were cautioned not to talk about their choices. They were then shown each of the 16 toys, one by one, and asked to circle which one of the two (or four) on their choice sheet they wanted as their reward for helping us. The toys, they were told, would be delivered to the school the same week.

After the choices had been made and the choice sheets collected, *E* explained what additional help the Ss were being asked to give. To do this he said:

You'll remember I said we were going to ask you to do something more for us this week. Well, what we want you to do is to indicate for us again just how much you like or dislike each of these 16 toys. You see a person's liking for toys like these may change from one week to the next. You might like a toy more, less, or about the same as you did last week and it's important for the toy manufacturers to know about this. So we are now going to ask each one of you to fill out a questionnaire just like the one you filled out last week.⁵

The toys were again held up, one by one, but this time somewhat more rapidly since they were already familiar to the Ss. Upon completion of the ratings the questionnaires were collected.

At this, the end of the experiment, the deception, design, and purpose were completely explained to the Ss. It was further explained that the experimenters were unable to afford a toy for each person but that instead the toys would be divided up among the eight classes participating, so that each class would receive four toys for use in the class room itself.⁶

Manipulation of Number of Alternatives. As indicated in the above instructions, the Ss in some classes were told they could choose between two of the toys, and their individual choice sheets designated the two

³ The authors wish to express their great appreciation to the following people for making this study possible: Charles T. St. Clair, Superintendent of North Haven Schools, Thomas P. Ryan, Principal at Temple School, Mildred A. Wakeley, Principal at Center School, and Edward L. White, Principal at Ridge Road School, and all of the sixth-grade teachers at these schools.

⁴ The authors would like to thank Mary Lee Brehm and Robert Ascheim for acting as helper at some of the experimental sessions.

⁵ The first class run was additionally asked to write out as much as they could remember of the characteristics of the choice objects. However, the Ss were unable to handle this task so no further groups were asked to do it.

⁶ While the school officials and teachers involved in this study reported no apparent ill effects, the authors were somewhat dissatisfied with their ability to reduce the Ss' residual tension. We now believe that a major deception of this type should be used at this age level only if the catharsis can be conducted individually or in small groups so the *E* can make sure of its success.

toys by name and number. Ss in other classes were told they could choose one from four toys that again were designated on their choice sheets by name and number.

Manipulation of Qualitative Similarity of Alternatives. The 16 toys rated and used as choice alternatives were chosen a priori to be relatively similar or dissimilar to each other. Two sets of four toys were picked so that there would be high similarity among the four in each set. The remaining eight toys were picked to be relatively dissimilar to any in the two sets of four, and to most of the other seven. The criterion of qualitative similarity was the similarity of activity that the toy would require, as indicated on the following classified list in which the toys appear in the same order as in the experiment:

- | | |
|--|--|
| 1. Swimming Fins | 5. Set of 52 Games |
| 2. Swimming Mask | 6. Set of 66 Games |
| 3. Snorkel | 7. Set of 52 Games
(slightly different from #5) |
| 4. Life Jacket
(All for water activities) | 8. Set of 58 Games
(All indoor table games) |
| 9. Indoor Bowling | 13. Aluminum Craft |
| 10. Ship Model | 14. Erector Set |
| 11. Copper Craft | 15. Archery Set |
| 12. Racing Game | 16. Metal Tapping Craft |

To create relative similarity, all of the alternatives offered were chosen from Numbers 1-4, or from 5-8. To create dissimilarity, all of the alternatives were chosen so they involved different activities, e.g., Numbers 1, 5, 14, and 16.

Questionnaires. The main rating questionnaire consisted of 16 rating scales, each numbered for one of the 16 toys. Each scale consisted of a horizontal line of 51 equally spaced dots with six equally spaced markers labeled: Don't like at all; Like just a little; Like fairly well; Like quite a lot; Like very much; Really like very, very much. Numerical values were given to the ratings by assigning values 0 through 5.0 to these labels, respectively, and scoring each answer to the nearest tenth of an interval (i.e., to the nearest dot on the scale). Ss were instructed to make a pencil check mark anywhere along the dotted line. There was a space for name and sex at the top of each questionnaire.

Design. Half of the eight classes were run in the Two Alternative condition, the other half, in the Four Alternative condition. Half of the Ss in each class were given Similar Alternatives, the other half, Dissimilar Alternatives.

The classes at any one school were run during a single morning on the same week day but one week apart for the two experimental sessions. Work at the three schools was conducted on three successive days.

Controls. While all Ss were given a choice and asked to fill out the questionnaires, only a portion of them yielded usable data. Many were excluded by requirements that had to be met to permit an appropriate test of the hypotheses. First, the choice alternatives for each S had to be liked, but not so much that an increase in liking would be impossible. Second, one alternative had to be initially more liked than any other so that its choice could be expected. Increased ratings of the chosen item are thus not likely to be simply a result of

TABLE 1
CHANGES IN LIKING OF CHOSEN AND INITIALLY NEXT MOST LIKED ALTERNATIVE

Number of Alternatives	Qualitative Character			
	Similar	N	Dissimilar	N
Two				
Chosen	0.07 ^a	29	0.20	21
Unchosen	-0.09	29	-0.65	21
Four				
Chosen	0.48	10	0.62	12
Unchosen	-0.27	10	-0.72	12

^a A positive number indicates increased Liking, a negative, decreased Liking.

normal (and random) changes in *actual* liking from the first questionnaire to the second. In addition, Ss who failed to choose the alternative initially marked as most liked were excluded because they gave unreliable or invalid ratings. Finally, in order to ensure that initially less liked alternatives were seriously considered as possible choices, initial ratings of these alternatives could not be much lower than the most liked alternative.

The stringency with which these requirements were applied was determined by the actual distribution of initial ratings and the need for a workable number of Ss in each of the four experimental conditions. The final criteria were: (a) the most liked alternative had to receive a rating between 2.9 and 4.6, or in other words, a moderate to high Liking score; (b) the next most attractive alternative could be no more than 1.1 scale points lower, but had to be at least 0.1 scale points (the least measureable difference) lower, than the first; (c) for four alternatives, the additional criterion was imposed that the least liked must be no more than 2.5 scale points below the most liked alternative.

RESULTS

Liking change scores were computed for all choice alternatives for the selected Ss. In order to compare the Four Alternative conditions with the Two, we have taken the chosen and the initially next most attractive alternative of the remaining three. Our hypotheses predict increased liking for the chosen alternative and decreased liking for the unchosen alternative, to the extent that dissonance is created by the choice. Thus we would expect greater change in liking in the Four than in the Two Alternative condition, and greater change in liking in the Dissimilar than in the Similar Alternative condition.

Table 1 presents the relevant data. Both hypotheses are supported. The chosen alternative in each of the four experimental conditions shows an increase in Liking, while the unchosen

TABLE 2
TOTAL DISSONANCE REDUCTION (CHANGES IN LIKING)
FOR COMBINED EXPERIMENTAL CONDITIONS

	Dissonance Reduction	N
Number of Alternatives		
Two	0.45	50
Four	1.08	22
Qualitative Character		
Similar	0.31	39
Dissimilar	1.03	33

alternative always shows a decrease in Liking. If we take the sum of changes in Liking for chosen and unchosen alternative as an index of total reduction in dissonance and combine conditions in order to arrive at over-all dissonance reduction scores for each of the two experimental manipulations, then we obtain the figures given in Table 2. Here it is clear that the Four Alternative condition produced greater dissonance reduction than did Two (p equals about .02),⁷ and the Dissimilar Alternative condition produced more dissonance reduction than did the Similar (p less than .01).⁸

The criteria used in selecting alternatives and Ss produced a small bias in the initial difference between the two alternatives in question such that the initial difference in the Four Alternative condition was less than that in the Two. Since the initial relative attractiveness of alternatives has been shown to determine amount of dissonance reduction (Ehrlich, Guttman, Schonbach, & Mills, 1957) it is necessary to investigate whether or not the present bias can alternatively account for the obtained results. When this is done by selecting out of the Four Alternative conditions those Ss for whom the initial difference was relatively large and comparing their data with those for Ss whose initial difference was small, the mean score for total dissonance reduction turns out to be the same regardless of size of

⁷ All p values reported were obtained by two-tail t tests.

⁸ The same computations for our unselected population yield support for the effect of the Similar-Dissimilar manipulation, but fail to show any effect of the number of alternatives. The very reason for selection of Ss was, of course, that data from these unselected Ss would fail to adequately test the hypotheses. Consequently, no conclusion should be drawn from the data for these unselected Ss.

initial difference. This small bias thus does not account for the obtained difference between the Two and Four Alternative conditions.

Changes in Liking for the Third and Fourth Alternatives. In the Four Alternative conditions, it is possible to further investigate the effects of Similar versus Dissimilar alternatives by comparing changes in Liking of the third and fourth alternatives. One might expect that such changes for these somewhat less attractive alternatives would follow the same pattern that held for changes in Liking of the previously considered unchosen alternative, but with smaller magnitude because of their lesser importance.

Mean change in Liking was therefore computed for the third and fourth alternatives in the Similar and Dissimilar conditions. However, since the initial rating of these alternatives tends to be below average, some positive change in rating may be expected solely from regression effects. In order to obtain a better estimate of Liking change it is necessary to correct the rating change by subtracting the estimated regression. Estimates of regression were obtained by computing the mean Liking change according to amount of initial rating for all toys not involved in the choice.⁹

The corrected Liking changes for the third and fourth alternatives, presented in Table 3, show an over-all increase in liking rather than the expected decrease. Furthermore, there is no consistent difference between the Similar and Dissimilar conditions. Our expectations for the third and fourth alternatives are clearly disconfirmed by these results.

⁹ To relate rating change to the initial rating, the Liking scale was divided into the following intervals: 0-0.5; 0.6-1.5; 1.6-2.5; 2.6-3.5; 3.6-4.5; 4.6-5.0. The mean Liking changes then obtained from all selected Ss for toys not involved in the choice and with initial ratings in each of the above categories were: +.5; +.5; +.1; -.1; -.3; and -.5, respectively. These corrections were then subtracted from each individual's Liking change scores according to his initial ratings of each of the alternatives.

Use of these corrections for the chosen and most-liked unchosen alternatives reduces slightly the obtained p values but in no way alters the pattern of dissonance reduction scores or conclusions that may be drawn from them. Thus with these corrections also, greater liking changes occur for dissimilar than similar and for four than two alternatives.

TABLE 3
CORRECTED LIKING CHANGES FOR THE THIRD
AND FOURTH ALTERNATIVES

Qualitative Character	Alternative		N
	Three	Four	
Similar	0.10	0.94	10
Dissimilar	0.58	0.21	12

DISCUSSION

The primary data from the present experiment support two derivations from Festinger's theory of cognitive dissonance: the magnitude of dissonance and consequent re-evaluation of choice alternatives is a direct function of (a) the number of choice alternatives and (b) the qualitative dissimilarity of the choice alternatives.

The effect of our manipulation of the number of alternatives may, however, be explained in a different way. It will be recalled that the magnitude of dissonance and consequent re-evaluation of choice alternatives is a function of the difference in initial liking for the alternatives: the smaller this initial difference, the greater the magnitude of dissonance after the choice. It follows that with other things held constant, the more liked are the unchosen alternatives, the greater the magnitude of dissonance. It is clear, then, that the Two and Four Alternative conditions must be equated on the initial liking of both the chosen and most liked unchosen alternatives. This was done, of course, as part of our procedure for selecting *Ss* whose choices would furnish a good test of the hypotheses. But the low reliability of our measure of liking means that where there are three unchosen alternatives, the one that was marked as initially most liked will not always be the one that was *actually* most liked. Thus the *actual* liking for the most liked unchosen alternative should tend to be higher where there are three unchosen alternatives than where there is only one. In addition, let us assume that, with liking for the magnitude of the chosen alternative held constant, the most liked unchosen alternative is a disproportionately more important determinant of dissonance than are the less liked alternatives. If this assumption is correct, then the differ-

ence in liking for the unchosen alternative implies that dissonance should have been greater in the Four than in the Two Alternative condition. It seems unlikely, however, that this bias could have been large enough to have produced the observed effects. While this alternative explanation is possible, it is not entirely plausible.

A more serious question may be raised in regard to the data for the third and fourth choice alternatives. For while the re-evaluation of the chosen alternative clearly reflects greater dissonance in the Four than in the Two Alternative condition (p equals .04), the third and fourth alternatives fail to show any sign of dissonance reduction. Indeed, the increase in liking for these alternatives should lead to *increased* dissonance. One possible explanation for this apparent increase is that our estimate of regression is too small. It may be, for example, that when a toy is made an alternative for choice, it tends, prior to the choice, to be re-evaluated more than a toy not made one of the alternatives. If the direction of such re-evaluations was random, our estimate of regression would be too small, thus accounting for the apparent increase in liking. It is also possible that such re-evaluation is not random, but rather is directed toward enhancing the value of all alternatives. The latter explanation would not only account for the increased liking of the third and fourth alternatives; it would also account for the fact that the chosen alternative increases in liking more than the initially most liked unchosen alternative decreases. While the phenomenon can thus be plausibly explained, it is of sufficient importance to the theory of dissonance to warrant separate and thorough investigation.

It is also of interest that the present experiment produced no visible interaction of the two independent variables in their determination of the magnitude of dissonance created by choice. Under certain conditions such an interaction would be expected. Thus if the qualitative similarity of alternatives is extreme (approaching identity) little or no dissonance would be created by choice no matter how many alternatives were involved. But from that point, increasing the dissimilarity should have more effect on dissonance where several alternatives are involved than where only two

are involved. Though we have no independent evidence from the present experiment, we may assume that the alternatives in the Similar condition were seen as being somewhat dissimilar. If this is the reason that no interaction effect was obtained, then a good test of the interaction hypothesis must employ a much more powerful manipulation of qualitative similarity than was used here.

SUMMARY

In order to test two derivations from Festinger's theory of psychological dissonance, sixth-grade school children were given systematically varied choices from among several attractive toys. About half of them were given a choice between two toys, and the remainder were given a choice of one from among four toys. From each of these two groups, about half was given a choice between qualitatively similar toys, while the remainder was given a choice between qualitatively dissimilar toys. Ss were asked, both before and after the choice,

to indicate on a 51-point graphic rating scale how much they liked each of the toys.

The two derivations tested were: the magnitude of dissonance and consequent amount of increased liking for the chosen alternative and decreased liking for the unchosen alternative increases with (a) the qualitative dissimilarity of alternatives, and (b) the number of alternatives. The first of these was confirmed; the second received somewhat questionable support. Changes in liking for the third and fourth unchosen alternatives failed to support our expectations. Two explanations for this failure were suggested.

REFERENCES

- BREHM, J. W. Postdecision changes in the desirability of alternatives. *J. abnorm. soc. Psychol.*, 1956, **52**, 384-389.
- EHRlich, DANUTA, GUTTMAN, I., SCHONBACH, P., & MILLS, J. Postdecision exposure to relevant information. *J. abnorm. soc. Psychol.*, 1957, **54**, 98-102.
- FESTINGER, L. *Theory of cognitive dissonance*. Evanston: Row, Peterson, 1957.

Received March 17, 1958.

INCREASING COGNITIVE DISSONANCE BY A *FAIT ACCOMPLI*¹

JACK W. BREHM²

Yale University

A RECENT theory by Festinger (1957) attempts to state the conditions under which discrepancies between cognitions produce "cognitive dissonance," a state of tension that motivates the individual to reduce or eliminate the discrepancies. One of the conditions that may be necessary for the creation of cognitive dissonance is that there be some element of volition. The theory states, for example, that a person who is completely forced to behave in a manner he would avoid if possible, experiences no dissonance. On the other hand, a *fait accompli*—i.e., an event outside of the person's control—might conceivably create dissonance if that same event would have led to the opposite behavior had it been predictable at a prior choice point. The purpose of this paper, then, is to report an experiment in which a *fait accompli* does appear to have increased cognitive dissonance.

The theoretical expectations may be stated succinctly. If a person is induced by the promise of a small prize to eat a disliked food, he should experience dissonance. If other things are held constant, the amount of dissonance can be increased by increasing the amount of the disliked food to be eaten. A person who is experiencing such dissonance should try, in proportion to its magnitude, to reduce it by convincing himself that he likes the food more than he originally thought.

If, then, the amount of expected eating is increased by a *fait accompli* after the initial decision to eat has occurred, and this is accompanied by increased liking, it may be concluded that a *fait accompli* can affect the magnitude of dissonance.

METHOD

In order to examine the effect of a *fait accompli*, eighth-grade students were offered attractive prizes for eating a small portion of a disliked vegetable.

¹ This study was conducted as part of the Yale Communication Research Program and was supported by a grant to Carl I. Hovland from Bell Telephone Laboratories. The author would like to thank Arthur R. Cohen, Carl I. Hovland, Leon Festinger, and Irving Sarneckoff for their comments on the manuscript.

² Now at Duke University.

About half of them were told, after eating the vegetable, that their parents would be informed which vegetable they ate. Liking for the vegetable was measured before and after this procedure.

The pre-experimental questionnaire. Approximately three weeks prior to the beginning of the experiment, questionnaires were given by teachers to all eighth-grade students in a New Haven Junior High School.³ Students were told that the questionnaire was sponsored by a nonprofit organization interested in people's eating habits. The questionnaire asked for each of 34 vegetables, "Considering everything about it (taste, smell, appearance, texture), to what extent do you like or dislike this vegetable?" Answers were indicated by making an X on any one of 61 dots running along a straight line and identified each 10 points by the following labels: like very much; like a lot; like a little; neither; dislike a little; dislike a lot; dislike very much. Subjects' responses were then scored from plus 3.0 for "like very much" to minus 3.0 for "dislike very much."

Experimental procedure. Subjects (Ss) were given passes during school hours to come one at a time to take part in a food-testing program. When an S arrived, the experimenter (E) introduced himself as being a representative of a fictitious consumer organization, whose business it was to test products and publish reports on them in order to help people in their shopping. It was explained that the organization was at present conducting a study of a number of vegetables by testing each for its food value in terms of vitamins and minerals and also by asking all kinds of people what they thought about them. E then said that what he wanted the S to do was answer some questions but only about one vegetable in order to get good information about that one. Before being given the main experimental questionnaire, the S was asked to answer some simple questions about his age, his favorite and most disliked foods, how often the vegetable picked for him was served in his home, and how often he ate it when it was served. It was also explained at this point that the vegetable in question had been picked at random, though actually it had been picked because it was one of the ones he most disliked (it was at least at -2.0 on the prequestionnaire). The S was then asked to fill out the main questionnaire carefully and accurately. To stress the anonymous nature of the questionnaires, the S was given a plain Manila envelope in which to put them when completed.

After the main questionnaire was completed, E said,

While we feel we get good information by asking a person to carefully fill out a questionnaire this way, we also know that we may get different kinds of

³ For their help in conducting this study, the author wishes to thank Louisa Willard, Psychologist for the New Haven Schools, Frank Carr, Principal of Troupe Junior High School, Mrs. Connolly, and the eighth-grade teachers, especially Mrs. Himmelfarb.

answers to these questions if a person actually tastes some of the vegetable before filling out the questionnaire. So we would like to have everyone eat a little of the vegetable and then answer some further questions about it. But we also realize that you might not like to eat just anything we put in front of you so the rest of this is completely up to you. You can stop right here without doing anything more at all, or you can eat a small dish of the vegetable and answer some more questions. However, to get as many people as possible to give us this further information, we are offering a prize to those who do go on. The prize consists of either two 45 or 78 rpm records of your own choice, or two tickets for movies of your own choice. You can do whichever you want. You can leave right now, or you can eat a small dish of — (name of vegetable) and get whichever of the two prizes you want, two phonograph records or two movie tickets.

If the *S* chose to quit he was thanked for his help and allowed to return to class. If he chose to continue, he was served a small dish of the vegetable, heated if appropriate. This constituted the *Low Consequence* experimental condition.

In the *High Consequence* condition, *Ss* were given exactly the same instructions. However, in order to create the impression that their choice to eat the vegetable would result in consequent further eating of it, they were casually given, when nearly finished eating, these further instructions:

Oh, I almost forgot to mention that one of the reports we plan to put out from this study will simply be a letter to the parents of each person who takes part, just indicating which vegetable that person ate.⁴ This information, which would likely have made many hesitate to go ahead with the eating if they had known about it, may be considered a *fail accompli*.

When the *S* had finished eating, he was asked to fill out again the main questionnaire. In making the request *E* mentioned that some people changed one way, some another, and some didn't change at all. He also stressed that the *S* should be as careful and accurate as possible and, finally, that the second questionnaire, like the first, was completely anonymous and to be slipped into the Manila envelope when completed.

When the *S* had finished the questionnaire, he was asked to choose which prize he wanted and told that the prizes would be sent in two or three weeks. He was also asked not to talk to other students about what happened in the testing procedure.⁵

When the study was completed, its true purpose was

fully explained to all *Ss*. The prizes were given as promised.

The experimental questionnaires. The first experimental questionnaire contained two important questions. The first of these asked *Ss* to indicate on a 51-point line with six equally spaced labels running from "every meal" to "never," how often the vegetable chosen for them was served at home. The second question, using an identical scale, asked how often they ate the vegetable when it was served. The main experimental questionnaire contained a question identical to that used on the pre-experimental questionnaire to measure liking for the vegetable.⁶

Design. In all, 20 *Ss* were run in the Low Consequence condition, 19 were run in the High Consequence condition. While an attempt was made to use the vegetables with equal frequency, this was only partly successful. Inspection of the dependent variables, however, reveals no large differences between vegetables.

RESULTS

Eating a small amount of a disliked vegetable should create some dissonance and consequent increase in liking in order to reduce the dissonance. If, in addition, the *fail accompli* (the High Consequence manipulation) increases dissonance, then persons exposed to it should experience greater dissonance and thus increase their liking for the vegetable more than those who are not exposed to it.

The mean change in Liking was therefore computed for *Ss* in each of the two experimental conditions. The prediction was confirmed in that the mean increased Liking for the High Consequence condition (3.76) was greater than the mean increased Liking for the Low Consequence condition (2.32). The difference between the conditions is significant at beyond the .01 level.⁷

While the instructions for the High Consequence condition do not obviously advocate liking the vegetable, it is conceivable they were interpreted in that way by the *Ss*. The differential increased liking for the vegetable between the two conditions might simply reflect differential "endorsement" of the vegetable by the "consumer organization." Thus our original explanation in terms of "dissonance reduction" requires further support to be cogent. We must therefore examine the manipulation and its effect more closely.

⁶ The main questionnaire also contained several questions included to gather data for a larger study of which this experiment was a part.

⁷ All *p* values in this paper were obtained by two-tailed *t* tests.

⁴ The idea for this manipulation came from some pretest *Ss* who spontaneously said, "Boy, would my mother like to know I'm eating this!"

⁵ Informal interviewing of *Ss* in the later part of the experiment indicated that there was a moderate amount of talking about the experiment. Most *Ss* had heard, before participating, what some of the vegetables were, that some people actually ate the vegetables, that prizes were being given away, and what the prizes were. None showed evidence, however, of being aware of the High Consequence Manipulation.

It will be remembered that right at the beginning of the experiment all Ss were asked to show how often their vegetable was served at home, and how often they ate it when it was served. If the "dissonance" explanation is correct, then the High Consequence manipulation, which depends on an implied increase in frequency of eating at home, should work best for Ss who indicated a large discrepancy between frequency of serving and frequency of eating. Ss who indicated little or no discrepancy between serving and eating should have experienced little or no increase in dissonance from the High Consequence manipulation. Therefore, those Ss who indicated a high discrepancy in frequency should show a greater increase in Liking from the Low to the High Consequence condition than should those who indicated a low discrepancy in frequency. If, on the other hand, the "endorsement" explanation is correct, there would be no reason to expect differential results between Ss indicating high, and Ss indicating low, discrepancy.

To make the indicated analysis, Ss within each experimental condition were split at the median into high and low discrepancy groups. For each of these groups the mean change in Liking was computed. It was found that for those individuals who indicated a low discrepancy in frequencies, the mean difference in Liking change (.67) between Low and High Consequence conditions was negligible, while for those who indicated a high discrepancy the difference (2.01) was significant at beyond the .01 level. These data thus support the conclusion that the High Consequence manipulation implies an increased frequency of eating at home and thereby increases the magnitude of dissonance and pressure to increase one's liking.

Further examination of the present data for evidence relevant to the "dissonance" explanation turned up an interesting phenomenon. Ss within each condition were divided, as close to the median as possible, according to their initial dislike for their vegetable. For each of these subgroups the final mean Liking score was computed.⁸ In this way it was found that

within both conditions those persons who indicated the greater initial dislike subsequently indicated greater Liking than did those with less initial dislike. Within the High Consequence condition, this difference is significant beyond the .01 level, and within the Low Consequence condition the difference yields a two-tailed *t* of about 1.0. It is improbable that persons with extreme initial dislike have less resistance to increasing their liking than do those with moderate initial dislike. We may therefore conclude that the greater was the individual's initial dislike, the greater was the experimentally produced pressure on him to increase his liking for the vegetable. This finding is consistent with dissonance theory, for as initial dislike increases, so should the dissonance from having to eat the vegetable.

DISCUSSION

Other research on the effects of a *fail accompli* has been imprecise as a result of the limitations of field studies (e.g., Clark, 1953), or, if done in the laboratory, has failed to produce experimentally adequate information (e.g., Brehm & Cohen, 1959). The present experiment, however, unequivocally demonstrates attitude change from a *fail accompli*, which is relevant to prior induction of unpleasant behavior.

If the differential increases in liking in the present experiment were indeed a result of differential amounts of dissonance, then the conceptual definition of dissonance has been further specified. The specification is not complete, however, for it may be that a *fail accompli* affects the magnitude of dissonance only when it is related to a previous choice. Further research is necessary to supply the answer to this and similar hunches.

A finding of perhaps equal theoretical interest was the relationship between initial strength of attitude and subsequent change. The more unfavorable was the initial attitude, the more favorable it tended to be at the end of the experiment. This relationship would be inconsistent with any theory of attitude change that ignores the size of discrepancy between positions held and advocated. It is, however, entirely consistent with dissonance theory. As the discrepancy between attitude and behavior increases, the amount of dissonance

⁸ Increased Liking from regression should increase with more extreme initial dislike and thus make the change scores unequal. For this reason, final Liking scores, which are not biased by regression or ceiling effects, were used for this analysis.

and consequent pressure to reduce it increase. Thus, the greater is the discrepancy between attitude and behavior, the greater will be the subsequent pressure to change the attitude. There is, of course, some resistance to changing one's attitude. The magnitude of this resistance might be expected to increase as the individual's position becomes more extreme. Under the conditions of the present experiment, the pressure to reduce dissonance apparently increased more rapidly with increasing discrepancy than did the resistance against change. This experiment centered around a relatively individualistic attitude (food preference). If, on the other hand, the initial attitude were supported by stronger, external sources of reality, it might not be overcome by the pressure to reduce dissonance. Again, further research is needed to explore the problem.

SUMMARY

This paper presents a study illustrating the effect of a *fait accompli* on cognitive dissonance. Eighth-grade boys and girls were induced by the promise of a prize to eat disliked vegetables. Some were additionally told, while eat-

ing, that their parents would learn what vegetable they had eaten, with the implication the Ss would have to eat more at home. A questionnaire measured liking for the vegetable before and after the experiment.

It was found that the *fait accompli* of implied further eating produced a greater increase in liking. This experiment therefore further specifies Festinger's concept of cognitive dissonance. In addition, it was found that persons initially least favorable toward the vegetable tended to be most favorable after the experiment. This result was seen as consistent with dissonance theory but not with some other theories of attitude change.

REFERENCES

- BREHM, J. W., & COHEN, A. R. Choice and chance relative deprivation as determinants of cognitive dissonance. *J. abnorm. soc. Psychol.*, 1959, **58**, 383-387.
- CLARK, K. B. Desegregation: An appraisal of the evidence. *J. soc. Issues*, 1953, **9**, 43-79.
- FESTINGER, L. *A theory of cognitive dissonance*. Evanston: Row, Peterson, 1957.

Received March 19, 1958.

CHOICE AND CHANCE RELATIVE DEPRIVATION AS DETERMINANTS OF COGNITIVE DISSONANCE¹

JACK W. BREHM² AND ARTHUR R. COHEN

Yale University

IT is not clear from Festinger's (1957) theory of cognitive dissonance whether events capable of producing dissonance under conditions of choice might fail to produce dissonance if the person could not exercise some choice in controlling its occurrence or his exposure to it. It has been shown, for example, that to the extent that a person is *forced* to do something he ordinarily would avoid, he experiences no dissonance (e.g., Cohen, Brehm, & Fleming, 1959). It would seem, then, that a chance event that is outside the person's control may fall in the class of phenomena that fail to produce dissonance. Similarly, a *fail accompli* is outside his control and may produce no dissonance. An experiment by Brehm (1959), however, demonstrates that a *fail accompli* can affect the magnitude of dissonance created by a prior choice. Thus, while the limits of conditions that create dissonance may be greater than those stipulated by Festinger, just where these limits lie is not yet known. The purpose of the present paper is to report an experiment designed to establish some of these limits.

Two questions will be examined in the present study: Can a chance event affect the magnitude of dissonance? Does the effect of such a chance event depend upon there having been a prior relevant choice? In order to clarify these issues, we must discuss two theoretical approaches to the setting in which the test was made.

The first is Merton's analysis (Merton & Kitt, 1950) of the relationship between "relative deprivation" and felt satisfaction. According to this view, the degree to which a person is satisfied with a given state of affairs may be determined not necessarily by his absolute level of achievement or benefit, but instead by how much he achieves or benefits *relative to those around him*. The greater is the relative

deprivation, the less will be the felt satisfaction with any arbitrary assignment or *fail accompli*.

The second theoretical approach is Festinger's analysis (1957; Festinger & Carlsmith, 1959) of forced compliance situations. This approach asserts that a person who is induced to behave in a way he ordinarily would avoid, experiences "dissonance"—i.e., psychological tension—which he then tries to reduce or eliminate. Other things being equal, the more unpleasant the induced behavior, the greater the magnitude of dissonance. Thus as relative deprivation increases, the magnitude of dissonance should increase. If a person is induced to perform an unpleasant behavior, thus creating dissonance, he can reduce or eliminate the dissonance by convincing himself that the behavior is really quite satisfactory after all (Brehm, 1957). His final attitude toward the event is therefore in part a direct function of the amount of dissonance created. The greater the relative deprivation, the greater is the amount of dissonance created and, consequently, the more positive his attitude toward the event.

With a setting of relative deprivation, then, we can test the points of theory at issue. By varying the amount of force used to make persons perform an unpleasant behavior, their feeling of choice about whether or not to perform it may be varied. After the person has committed himself to performing the unpleasant behavior, he may be told that due to a chance factor, he will be deprived relative to other persons who perform the same behavior. The amount of relative deprivation can be varied. We might then expect that under low choice conditions, increasing relative deprivation would result in decreasing satisfaction. Our major hypothesis thus is that under greater but not under lesser degrees of choice, increasing relative deprivation tends to result in increasing satisfaction. Thus if chance relative deprivation affects dissonance only under conditions of prior choice, there should be an interaction between amount of choice and

¹ This experiment was performed as part of the Yale Communications Research Program. It was supported by grants to Carl I. Hovland from the Bell Telephone Laboratories and from the Rockefeller Foundation.

² Now at Duke University.

amount of relative deprivation in determining the amount of subsequent satisfaction.

METHOD

Subjects (Ss) were told that they must take part in a disagreeable and time-consuming experiment. Approximately half of the Ss were given a high opportunity, and half a low opportunity, to get themselves excused; within these groupings, half were placed in a situation of high relative deprivation and half in one of low relative deprivation. Immediately thereafter a questionnaire was administered to measure the Ss' attitudes toward participating in the experiment and to check the success of the experimental manipulations.

Subjects

Five sections of the introductory psychology course at Yale University were used. The Ns were: 24, 24, 28, 29, and 29. Ss were approached in their classrooms during a regularly scheduled class period.

Procedure

The experimenter (E) was introduced as a member of the psychology faculty who was on a departmental committee concerned with student participation in psychological research. E explained that a new project had just been set up in which all students in introductory psychology classes were to participate. He said that unfortunately, this particular project wasn't going to be very interesting. He described the task as three solid hours of copying random numbers in order to establish norms for other research. He emphasized that it would be boring, tedious, no challenge, since they must not try to be extra fast or accurate in their copying, and not a very good opportunity to learn about psychological research.

Manipulation of Choice. Ss were then given schedule sheets on which to indicate their free hours. In the *Low Choice* condition, these sheets were collected without mention of any way for the Ss to be excused. In the *High Choice* condition, Ss were told that though it was desirable that everyone participate, a person could be excused if necessary. E told them that to be excused, they must indicate on a separate form attached to their schedule sheets their desire for an appointment with the Psychology Department's Director of Undergraduate Studies concerning their excuse. Two classes were given the *High Choice* manipulation, three were given the *Low Choice* manipulation.

Manipulation of Relative Deprivation. When the schedule sheets had been collected, the Ss were asked to evaluate the general issue of participation in research. Attached to each questionnaire was a page of "evaluation sheet instructions." These instructions, intended to induce the conditions of relative deprivation, said:

In this research project you are to copy random numbers in an "average" way. You must not try to be unusually accurate or fast. A large body of evidence shows that rewards don't make much difference in this kind of research, and they probably won't be too important here either. Nevertheless, looking at the effects of rewards might not be a bad idea. Thus,

while most persons will receive \$_____ for their three hours of copying, a few, randomly chosen people will receive nothing. The following check mark indicates whether or not you personally will receive payment.

Will receive payment _____
Will not receive payment _____ X

In the *Low Relative Deprivation* condition, the Ss were told that "most persons will receive \$1.00" for their copying. In the *High Relative Deprivation* condition the Ss were told that "most persons will receive \$10.00" for their participation. Half of each class were assigned to the high condition and half to the low condition. In both conditions, Ss were told that they would receive nothing. It should be noted that each person was told he was relatively deprived by *chance*.

In manipulating deprivation, it was necessary to keep the differential reward plausible, though not too plausible. To the degree that there were good reasons for others getting money, dissonance would be reduced. On the other hand, an entirely implausible reason would militate against the seriousness of the entire enterprise.

After reading the above "information," the Ss completed the questionnaire. The central dependent variable of the present study concerns the degree of satisfaction experienced by the Ss at the prospect of participation in the disagreeable research project. The degree of satisfaction was measured by the Ss' ratings on a nine-point scale which asked them the following: "Considering what you will get out of the experiment (interest, enjoyment, learning, etc.) how satisfied are you with having been assigned to this research?" Checks on the experimental inductions were also made. For the choice variable, the Ss were asked on a nine-point scale, "How difficult do you think it would be for you to get out of this experiment?" As a check on the relative deprivation inductions, the Ss were asked whether or not they would receive anything for their participation and whether or not they felt that they should receive anything.

The questionnaires were then collected and the Ss were apprised of the nature of the experiment. They were assured that they did not have to take part in the "research," and through an illustrative talk on experimental design, whatever annoyance or anxiety they had was presumably dissipated.

RESULTS

Effectiveness of the Experimental Manipulations

The data indicate that the manipulations of relative deprivation were entirely effective in creating perceptions of the differing amounts of money to be earned by others in the same class.

The variable of choice, however, presents another picture. In examining the Ss' perceptions of the degree of choice they had in submitting to the *fait accompli*, it was found that more variance in choice existed within experimental conditions than between them. In

TABLE 1
MEAN RATINGS OF SATISFACTION FOR EACH
EXPERIMENTAL SUBGROUP^a

	Perceived Degree of Choice	
	Low	High
Replication #1 (N = 29)		
Low Rel. Dep.	5.19 ^a [9]	4.00 [6]
High Rel. Dep.	3.82 [8]	5.50 [6]
Replication #2 (N = 24)		
Low Rel. Dep.	2.57 [7]	2.12 [5]
High Rel. Dep.	2.06 [8]	3.10 [4]
Replication #3 (N = 28)		
Low Rel. Dep.	4.81 [7]	3.04 [7]
High Rel. Dep.	2.58 [8]	3.78 [6]
Replication #4 (N = 24)		
Low Rel. Dep.	3.77 [3]	2.11 [10]
High Rel. Dep.	1.58 [6]	2.85 [5]
Replication #5 (N = 29)		
Low Rel. Dep.	4.94 [7]	3.07 [8]
High Rel. Dep.	3.84 [9]	2.34 [5]

^a Small numbers indicate low satisfaction, large numbers, high.

effect, then, the manipulation of choice failed to determine the amount of choice felt by the Ss. Our analyses therefore employ the degree of *perceived* choice to distinguish between those who experienced a high degree of choice and those who felt that there was little choice in submitting to the disagreeable *fait accompli*. Since the variances in choice differed from class to class, each class was dichotomized on choice as near as possible to the median, and each such subgroup further divided into the manipulated conditions of High and Low Relative deprivation. Thus we have five independent replications of an experiment where perceived choice varies and where relative deprivation varies but where all Ss expect to participate in a disagreeable research project.³

Effects of Choice and Relative Deprivation on Satisfaction

Ratings of the Ss' attitude toward the research assignment constitute the data of the present experiment. In Table 1 are presented the mean satisfaction ratings for each of the four conditions for all five replications. Corresponding Ns are presented in parentheses.

³ There are no data on expectation for participation in the conditions where Low Choice was manipulated; we must therefore assume that all Ss perceived themselves to be committed.

TABLE 2
INTERACTION DIFFERENCES BETWEEN HIGH AND
LOW RELATIVE DEPRIVATION CONDITIONS
FOR EACH CHOICE CONDITION

Replication	Low Choice	High Choice	Interaction Difference	p Value*
1	1.37	-1.50	2.87	.04
2	.51	-.98	1.49	ns
3	2.23	-.74	2.97	.07
4	2.19	-.74	2.93	.06
5	1.10	.73	.37	ns

* All p values are from two-tailed t tests.

The results are highly consistent. Except for a reversal of means in the High Choice condition in Replication 5, all the data support the assumption of an interaction between degree of choice and degree of deprivation in determining attitudes toward a prospective event. These results are summarized in Table 2 where the differences between deprivation conditions within the choice conditions and the interaction differences are presented. In the first two columns of Table 2, a positive figure indicates that those Ss in the Low Relative Deprivation condition were more positive in their attitudes, a negative figure indicates that the High Relative Deprivation Ss were more positive.

Table 2 shows that in all five replications the interaction differences are in the same direction. In one of the two cases where the difference is statistically unreliable (Replication 5) it is due to a reversal in the High Choice condition. In general, however, we may conclude that attitudes toward the unpleasant *fait accompli* are strikingly different depending upon degree of perceived choice and degree of relative deprivation experienced after commitment.

DISCUSSION

If we can assume that the present results would have been obtained had the manipulation of choice been successful, then our major hypothesis appears to have been supported. When an individual feels he has little prior choice in whether or not he is committed to an event, a direct effect of deprivation occurs: increasing relative deprivation results in decreasing satisfaction. If an individual feels he has at least some prior choice, increasing rela-

tive deprivation results in increased dissonance and consequently in increased satisfaction.

Since Ss in the present experiment thought they were relatively deprived as the result of chance, certain kinds of events not under the person's control appear capable of producing dissonance. At the same time, a limit to these conditions is found. The data indicate that only when the person feels he has had some prior choice can a chance event affect the magnitude of his dissonance.

A second point of theoretical interest concerns the qualification of Merton's concept of relative deprivation. In the present data his analysis may apply only when the individual has had no choice concerning his assignment.

It should be noted that although relative deprivation was induced in the present experiment, any negative chance event introduced after commitment in a situation of High Choice would have the same theoretical status. Thus variations in the degree of any type of punishment might be expected to create similar variations in dissonance and consequent attitude change. In addition, as far as dissonance theory is concerned, differential relative deprivation ought to produce differential dissonance even in cases in which persons receive some rewards for engaging in the negative behavior. In the present experiment the Ss varied in the degree to which they were punished relative to others. It would also seem that under conditions where forced compliance has been elicited to a disagreeable behavior, variations in the degree to which others are punished relative to the individual and variations in the degree to which he is rewarded relative to others should also produce variations in dissonance. In all instances, the discrepancy between his cognitions supporting his behavior and those negating it, would vary with degree of relative deprivation. The emphasis on relative deprivation highlights a view of motivation for compliance as not absolute but as relative to the situation of others around the individual. And it would follow that the more important or similar or salient those others, then the greater would be the effect on dissonance of variations in relative deprivation.

Unexplained Effects. While our original

hypotheses account for some of the obtained effects, they do not account for all. We did not expect that persons in the condition of Low Relative Deprivation would tend to be less satisfied as their opportunity to be excused increased. There are several possible explanations of this apparent effect, of which we will mention one that seems plausible to us.

This explanation assumes that as the magnitude of force to perform an unpleasant behavior increases, a person tends less and less to try to reduce his dissonance by attempts to avoid its performance. In other words, the greater the force to comply, the more dissonance tends to be reduced by changes in attitude rather than by contemplation of avoiding compliance. Persons in a situation of Low Choice, compared to those in one of High Choice, should be more likely to reduce their dissonance by increases in satisfaction. In general, therefore, those confronted with High Choice might tend to show less satisfaction. This result would not, of course, alter the fact that more dissonance may be created in the condition of High Choice than in that of Low Choice, so that the obtained interaction may still be expected to occur.

Implications for Further Research. While the present data suggest that a chance event can affect the magnitude of dissonance if a person has made a prior relevant choice, this interpretation is somewhat equivocal because of the failure of the choice manipulation. A true experimental demonstration that a chance event can affect dissonance is still needed.

A second ambiguity of interpretation centers on the fact that persons in the condition of Low Relative Deprivation tended to be less satisfied as their opportunity to be excused increased. This paradoxical phenomenon deserves thorough investigation. One possible line of research might follow the explanation suggested above: that increasing the force to comply decreases one's ability to reduce dissonance by contemplation of escape from compliance. One might expect, for example, that this would occur where coercion is used, as in the present example, but would not occur where the inducing force was a reward. Other possible explanations must also be systematically investigated experimentally.

SUMMARY

This experiment was designed to explore some of the limiting conditions under which cognitive dissonance and consequent attitude change are produced. Two questions were asked: (a) can a chance event affect the magnitude of dissonance and (b) does the effect of such a chance event depend upon there having been a prior choice in commitment to an event?

In order to explore these hypotheses, perceived choice in submitting to an unpleasant event and a chance relative deprivation after commitment was varied. Five independent replications were carried out in five introductory psychology sections. The Ss were asked to and did commit themselves to participation in a disagreeable and time-consuming project. They varied in their perception of how much choice they had in getting out of it if they wanted to, and in the degree to which relative deprivation in connection with their participation was introduced after commitment.

The results indicate an interaction such that under conditions of low perceived choice, the more relative deprivation Ss suffer, the more *negative* are their attitudes toward the *fait accompli*, while under conditions of high choice, the more relative deprivation they suffer, the

more *positive* are their attitudes toward the *fait accompli*. Thus, a chance event may be seen to affect the magnitude of dissonance and consequent attitude change but only under conditions of high choice.

The results were discussed in terms of the bearing of dissonance-producing chance events and relative reward and punishment on dissonance theory. Alternative explanations for the results were also offered.

REFERENCES

- BREHM, J. W. Behavioral compliance and attitude change. Depart. of Psychol., Yale Univer., 1957. (Mimeo.)
- BREHM, J. W. Increasing cognitive dissonance by a *fait-accomplis*. *J. abnorm. soc. Psychol.*, 1959, **58**, 379-381.
- COHEN, A. R., BREHM, J. W., & FLEMING, W. H. Attitude change and justification for compliance. *J. abnorm. soc. Psychol.*, 1958, **56**, 276-278.
- FESTINGER, L. *A theory of cognitive dissonance*. Evanston: Row Peterson, 1957.
- FESTINGER, L., & CARLSMITH, J. H. Cognitive consequences of forced compliance. *J. abnorm. soc. Psychol.*, 1959, **58**, 203-210.
- MERTON, R. K., & KITT, ALICE. Contributions to the theory of reference group behavior. In R. K. Merton & P. F. Lazarsfeld (Eds.), *Studies in the American soldier*. Glencoe: Free Press, 1950.

Received March 19, 1958.

ATTITUDINAL EFFECTS OF CHOICE IN EXPOSURE TO COUNTERPROPAGANDA¹

ARTHUR R. COHEN, HERBERT I. TERRY, AND CHARLES B. JONES

Yale University

IN A series of experiments stimulated by Festinger's (1957) theory of cognitive dissonance it has been found that when a person is induced to take a behavioral stand contrary to his cognitions he may reduce the dissonance or tension between his behavior and cognitions by making those cognitions (i.e., opinions) consistent with the behavior (Brehm: 1957a, 1957b; Cohen, Brehm, & Fleming, 1958; Festinger & Carlsmith, 1959). The theory also states that the introduction of information or opinions discrepant from the individual's own opinion also produces dissonance and consequent pressure to reduce that dissonance. However, it is not entirely clear from the theory whether simple exposure to discrepant information is enough to produce dissonance and consequent attitude change in line with the information, or whether or not a more definite behavioral stand on the part of the individual to expose himself to such information is necessary. Research (Brehm, 1959; Brehm & Cohen, 1959) has shown that chance events may produce cognitive dissonance and consequent attitude change without the person's taking a stand relative to that particular event. It appears, though, that at least *some prior decision* in which the person commits himself to the general situation is necessary. The present experiment is designed to study further the effects of choice on cognitive dissonance.

The theory states that under certain conditions exposure to contrary information produces dissonance between the person's prior cognitions and his new cognitions. Everything else being equal, we might expect that the greater the difference between the person's opinion and the new information, the greater the dissonance produced. The dissonance produced by exposure to new information may be reduced by misperception, misinterpretation, rejecting those who disagree, changing one's own opinion, influencing others, and so forth (Festinger, 1957). However, dissonance

reduction through change in opinion to fit the new information should be most likely to result when the person has in some active way chosen to expose himself to the information. Thus, under conditions where he has had little choice in exposure, a direct effect of persuasion might be expected to occur: the greater the discrepancy between his prior opinion and the new information, the greater his resistance to changing his attitudes in line with the new information. Where the individual has chosen to expose himself to counterpropaganda, greatest dissonance should be expected where his own opinions are most different from the new information (Brehm, 1959). The discrepancy between his own opinion and the behavioral stand involved in his decision to entertain counterpropaganda can be resolved by changing his opinions toward the new information. Our major hypothesis, thus, is: *Under greater but not under lesser degrees of choice, increasing the discrepancy between a person's initial opinion and new information confronting him gives rise to increasing dissonance and consequent attitude change.* Thus, if the difference between the individual's initial opinion and the new information affects dissonance only under conditions of prior choice, there should be an interaction between amount of choice and amount of discrepancy in determining the degree to which attitude change occurs.

METHOD

Subjects (Ss) were asked to take part in a "survey." They were first given a prequestionnaire on the attitude dimension, then given the major experimental manipulations. Half were given a high opportunity to avoid hearing counter-propaganda, and half a low opportunity. They were then exposed to the new information. Immediately thereafter a questionnaire was administered to measure the Ss' attitudes on the same dimension and to check the success of the experimental inductions.

Subjects

Seventy-one freshmen at Yale University comprised the sample. They were approached individually in their dormitories.

¹ This report was prepared as part of the Yale Communications Research Program.

Procedure

The experimenter introduced himself to each *S* as someone taking a survey for the "Department of Human Relations." The survey concerned the advisability of young men marrying before the age of 23 (about two years after graduation). The issue was a then current topic of discussion in the campus newspaper and among the undergraduates, and was thus assumed to be salient.

Ss first filled out an a priori six-point scale as to how they felt about "men marrying by the age of 23." The scale ran from "agree very strongly" to "disagree very strongly." When the *S* had completed the questionnaire, *E* noted which side of the attitude continuum the *S* was on and chose the communication that opposed the *S's* opinion. In all cases *Ss* received new information contrary to their initial opinions.

Manipulation of Choice. When the opinion measures had been collected *E* said to all *Ss*, "I have here some additional information on early marriage collected by the Department of Human Relations from the results of previous surveys, which I would like to read to you now. It does, however, uphold a point of view quite contrary to the one you have taken." In the *High Choice condition*, the *Ss* were told: "I'll leave it entirely up to you whether or not I read it to you. Would you like to hear it?" If *S* said yes, he was asked, "Are you sure?" If he said yes again, he was asked again if he was sure. If he still agreed, *E* then read the appropriate argument to him. In the *Low Choice condition*, *Ss* were merely told that *E* was going to read the contrary information to them.²

The Information. The issue of early marriage was chosen because of its salience and because it was relatively easy to construct arguments for each side of the issue. Both arguments contained some facts and some half-truths. In order to equate the strength of the opposing communications, arguments used on one side were merely negated and used on the other side. Evidence on the success with which arguments were equated will be presented later.

Discrepancy Between Initial Opinion and New Information. This variable was not manipulated experimentally but was controlled through the initial measures of the *Ss'* opinions. *Ss* who were extreme on the attitude scale on either one side of the scale or the other were assumed to represent the condition of *High Discrepancy*; *Ss* whose responses fell in the middle of the attitude scale were assumed to represent the condition of *Low*

Discrepancy. With arguments on a given side constant, the more extreme the individual is in his attitude, the greater the discrepancy between his opinion and the new contrary information. In addition to noting which side of the issue *S* was on before introducing the information, *E* also noted the extremity of his opinion in order to assure equal and random assignment into the discrepancy conditions.

Postmeasures. After hearing the new information, *Ss* filled out a postquestionnaire. They were told, "Now that you have heard this additional information on early marriage, compiled from the results of previous surveys, we would like to get some of your reactions to it. You may possibly want to evaluate the situation in the light of this information. So would you please fill out this questionnaire." *Ss* filled out an opinion scale identical with the premeasure and also an item checking on the choice manipulation. On the latter they were asked, "Considering that you agreed to participate in this research, how much choice do you feel you had in whether or not you were given the additional information on early marriages?" The scale used was a six-point scale which ranged from "complete choice" to "no choice at all."

Ss were then told in some detail about the nature and true purpose of the experiment and cautioned not to speak to others about it for an appropriate time.

RESULTS

Effectiveness of the Experimental Manipulations

The data indicate that the choice manipulations were entirely effective in creating perceptions of differing degrees of choice in exposure to the new information. The choice conditions show extreme differences in the expected direction: the mean for the *High Choice condition* is 1.96; for the *Low Choice condition*, 4.36. This difference is significant beyond the .001 level ($t = 9.21$).³

The data also show that the information given to the *Ss* was fairly convincing and that the opposing arguments were of approximately equal strength. When the change in opinion from before to after measurement is taken for the entire population, the average change is .69, which is significantly different from no change ($p < .05$). Furthermore, the change in opinion among those *Ss* who were pro and received con arguments is no different from the change among those who were con and received pro arguments. The average change among the former is .74, among the latter .68. Thus, we may safely assume that differences between our experimental subgroups need not be due to

² Since all *Ss* had already filled out a prequestionnaire, practically all *Ss* assigned to the *High Choice condition* agreed to hear the opposing arguments: only three *Ss* (with *High opinion discrepancy*) refused to entertain the counterpropaganda. These *Ss* were eliminated from the experiment. However, the deletion of these *Ss* is not enough to seriously bias the results. Furthermore, the fact that three *Ss* felt enough choice to refuse indicates that constraints against not listening to the opposing arguments were not overly great and that there was in fact high choice in this condition. Further evidence on the effectiveness of the choice manipulation will be given with the results.

³ All statistical tests reported here are two-tailed tests.

TABLE 1
MEAN OPINION CHANGE FOR EACH
EXPERIMENTAL SUBGROUP

Discrepancy Between Opinion and Information	Degree of Choice	
	Low	High
Low	.76 [19]	.63 [20]
High	.32 [17]	1.17 [15]

differences in the strength of the communications.

Effects of Choice and Discrepancy on Attitude Change

The change in *Ss'* opinions toward the marriage issue from before to after exposure to new counterinformation constitute the data of the present experiment. In Table 1 are presented the opinion change scores for each of the experimental subgroups. Corresponding *Ns* are in parentheses; the higher the figure, the more the change.

In Table 1 it can be seen that the data support the assumption of an interaction between degree of choice in exposure and degree of opinion discrepancy in determining attitude change toward new discrepant information. The interaction *t* equals 2.49, $p < .02$.

DISCUSSION

In the present experiment, the discrepancy between initial opinion and new information was not manipulated experimentally but was allowed to vary with the *Ss'* own opinion preferences. This procedure could allow some ambiguity of interpretation if self-selection into the choice conditions had been permitted or if initial opinions were highly correlated with attitude change. But since degree of choice was experimentally manipulated and there are no reliable over-all differences in opinion change between the High and Low Discrepancy conditions, we can assume that our major hypothesis has been supported. Thus, when a person feels that he has little prior choice in exposure to information counter to his opinion, increasing the discrepancy between his opinion and the information results in decreasing attitude change. If he feels he had at least some prior choice, increasing the discrepancy results in increased dissonance and consequent attitude change.

Under conditions of Low Choice, the relationship between discrepancy and attitude change may be conceived of as one of direct persuasion. The more extreme a person is in his own opinion the more he may be expected to resist counterarguments. Because he has had no choice in exposure, there is little dissonance between his cognitions (initial opinion) and behavior. What dissonance there remains between his opinion and the opposing information he can reduce in ways other than attitude change. However, when a person has chosen to expose himself to the countercommunication, dissonance is produced between his cognitions (initial opinion) and behavior (choice of listening to discrepant information). In addition, the more discrepant the information, the more dissonance should be produced. In a situation in which he has chosen to take a stand contrary to his cognitions, the dissonance can be reduced by making those cognitions consistent with the behavior (i.e., change his attitude to conform with the new information). And we would expect, the more the dissonance, the greater the pressure to reduce it and, therefore, the greater the consequent attitude change.

Though the original hypothesis accounts for most of the obtained effects, it may not account for all. It was not expected that in the Low Discrepancy condition, choice would make little difference in attitude change. An alternative explanation would assume that where no potentiality for dissonance exists as a result of the small discrepancy between initial opinion and new information, choice in exposure is not seen as a stand sharply etched against one's own cognitions. As the discrepancy increases, the decision to listen to counterinformation is seen as a behavioral commitment quite contrary to one's own position, resulting in increased dissonance and consequent attitude change. A more thorough exploration of this issue through the experimental manipulation of opinion discrepancy is still necessary.

Nevertheless, it may be said that exposure to counterinformation per se may not be enough to produce dissonance, but may only mobilize the usual methods of dealing with attempts at persuasion, depending upon how different from his own position the person sees the new material to be. In situations in which

he has actively chosen to expose himself to discrepant information, however, dissonance may result and lead to reduction through attitude change in line with the new information.

SUMMARY

This experiment explored some of the conditions under which exposure to information discrepant from one's own opinion produces cognitive dissonance and consequent attitude change. It was assumed that chance exposure to such information would produce little dissonance, but that the person's choice to expose himself to such information would produce high dissonance and consequent attitude change, the greater the discrepancy the greater the dissonance.

To test this hypothesis, choice in exposure to contrary information was varied, and the person's opinion on a salient attitude dimension was measured. Ss were approached individually and a premeasure of their attitude was taken. Half the Ss who were extreme in their opinions and half the Ss who were moderate were simply exposed to information counter to their opinions; the other half of these groups were given a choice in exposing themselves to such information. After exposure, their opinions were measured again.

The results indicate an interaction such that under conditions of *Low Choice*, a direct persua-

sion effect occurred: Ss' attitudes toward the issue changed more in line with the new information, the *less* the discrepancy between the information and their initial opinion. However, under conditions of *High Choice*, expectations from dissonance theory were fulfilled: Ss' attitudes changed more, the *more* the discrepancy between the information and their initial opinion. The results were discussed in terms of the importance of prior choice in exposure in creating dissonance.

REFERENCES

- BREHM, J. W. Forced verbal compliance and attitude change. *Dept. of Psychol., Yale Univer.*, 1957. (Mimeo.) (a)
- BREHM, J. W. Behavioral compliance and attitude change. *Dept. of Psychol., Yale Univer.*, 1957. (Mimeo.) (b)
- BREHM, J. W. Increasing cognitive dissonance by a *fait accompli*. *J. abnorm. soc. Psychol.*, 1959, **58**, 379-382.
- BREHM, J. W., & COHEN, A. R. Choice and chance relative deprivation as determinants of cognitive dissonance. *J. abnorm. soc. Psychol.*, 1959, **58**, 383-387.
- COHEN, A. R., BREHM, J. W., & FLEMING, W. H. Attitude change and justification for compliance. *J. abnorm. and soc. Psychol.*, 1958, **56**, 276-278.
- FESTINGER, L. *A theory of cognitive dissonance*. Evanston: Row Peterson, 1957.
- FESTINGER, L., & CARLSMITH, J. H. Cognitive consequences of forced compliance. *J. abnorm. and soc. Psychol.*, 1958, **58**, 203-210.

Received March 20, 1958.

EVALUATION OF THERAPY BY SELF-REPORT: A PARADOX¹

JANE LOEVINGER AND ABEL OSSORIO

Jewish Hospital, St. Louis, and St. Louis State Hospital

ROSE (1956) has recorded some data which suggest that soldiers whose responses on personality questionnaires are "normal" and indicate "high morale" are a little more likely to break down in battle than those who give more aberrant responses. Here is a paradox to which we should not blind ourselves by clinging to an arbitrary definition of normality. The idea of "normality" or "good adjustment" is a complex one that may contain mutually contradictory implications. Is the purpose of therapy always to increase "normality" or "good adjustment?"

Measurement and description of personality have at various times been stranded on two opposite shores of the main stream. On the one side are those who get lost in the particularities of the process they set out to describe: Here we find Erikson's (1950) seven stages in the normal process of ego development and Rogers' (1958) seven-stage description of the process of psychotherapy. However illuminating as descriptions, these intricate schemes do not lend themselves to psychometric assay.

On the other bank are the testers with their abstractions and oversimplifications: Everyone must be more or less of one thing, like "adjusted," or somewhere between a pair of poles, like "authoritarian-democratic." If one superposes Erikson's portrait of normal development, Rogers' portrait of psychotherapy, and the results of objective personality measurement, especially those pertaining to authoritarianism and response stereotypy, what emerges as from a composite photograph is a variable whose core is *ability to conceptualize oneself* as a psychological person.

Let us look first at this trait as an aspect of normal development. During the second or third year, the child first learns that he has impulses and that acting on them is in some cases forbidden and in other cases must be

deferred. That this time is the point of origin of the moral sense has often been emphasized. Here we emphasize instead that it is the point of origin of one's conception of his inner life. To conceive of one's impulses is to achieve some measure of control over them. One of the major tasks of childhood is extension and elaboration of the concepts of impulse and control.

By early adolescence there is normally established a strict moral sense, at least on the verbal level. The typical adolescent is rigid, prone to think in stereotypes, intolerant of deviations, punitive, and antipsychological—in short, what has been called an authoritarian personality (Erikson, 1950; Remmers & Radler, 1957). His self-conception is conventional and expressed in "socially desirable" terms.

During the college years, particularly in sophisticated Eastern schools, there is a decline in authoritarian tendencies and a shift to a liberal and tolerant view of others and a more sophisticated, differentiated, self-critical, and realistic self-concept (Christie, Havel, & Seidenberg, 1958; Cohn, 1957; Sanford, 1956). Where the adolescent's standards are typically expressed, "The right way is . . ." the more mature person's standards are more likely to be expressed, "I am the kind of person who . . ." In terms of the growth of the capacity for self-conception, we express it this way: The infant has no distance from his impulses; the adolescent has distance from impulse but not from ego; the mature person has capacity for distance from ego as well as impulse.

If we project these three stages onto an adult screen, we see that they correspond to meaningful individual differences. There are older children and even adults who have a degree of impulse control appropriate only to young children. Redl and Wineman (1957) have portrayed preadolescent children of this type. As adults they turn up as delinquents or psychopaths or disorganized personalities or as parents of delinquent children. They

¹ Preparation of this paper was supported in part by a research grant, M-1213, from The National Institute of Mental Health, Public Health Service.

come to social agencies and guidance clinics as involuntary referrals. However miserable their troubles, they "don't know why they were sent," and have nothing to say. They are unable to verbalize about themselves or even about their way of life. They come to an alcoholic clinic asking, "Where do I get the shots?" Their only conception of trouble is organic. It may take many weeks to bring such adults to the point of understanding what psychotherapy is about.

The differences between conventional-authoritarian adults and liberal-sophisticated adults are those which have been delineated by over a decade of research with the California F scale and similar tests. The great mass of the population is not much further along the scale of ego development than the typical adolescent. At the conclusion of the California study of the authoritarian personality, the authors came to realize that the psychopathic authoritarian and the conventional authoritarian are distinct types (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950, p. 972). They aimed to study the pathological type but ended by studying chiefly the conventional type. The pathological authoritarian appears to be drawn from what we have called the lowest level of ego development, the conventional authoritarian from our middle level. However, connotations of the term authoritarianism are too harsh, too pathological, and too gratuitously political for the present context; conventionality or even Philistinism is closer to characterizing our middle level.

If we define normality in statistical terms, the conventional person with a stereotyped self-concept is normal, whereas the more mature person with a more realistic self-concept is less normal because less numerous. On the other hand, if we take success in life as the criterion of good adjustment, clearly those who remain at the lowest level of ego development beyond infancy will be poorly adjusted in most instances. But there is probably little difference in life success between the conventional and the intellectually sophisticated. Elevation of the Philistine to places of power is a well-known hazard of democracy.

The conventional personality typically deals with unacceptable impulses by repression

and denial. This is part of what the California group called "anti-intracception." Attempts to disseminate publicly the findings of psychology, particularly when they undermine the average person's mental habits, are likely to strengthen his antipsychological defenses. He is by no means exempt from psychological illness, but he is debarred from utilizing the services of psychotherapists by his antipsychological defenses. If he comes into psychotherapy, it will often be because of medical referral for psychosomatic illness or because of an acutely disabling condition (Hollingshead & Redlich, 1958, Ch. 6).

To come voluntarily for psychotherapy for a mildly disabling condition already implies some sophistication in one's conceptualization of his inner life, a capacity for detachment and self-criticism. In addition, all forms of psychotherapy operate by means of increasing one's capacity for realistic self-conception. At this point we are tracing the connection between our concept of ego development and Rogers' concept of the process of therapy. While we emphasize the cognitive aspect, and therefore speak of increased capacity for distance from ego in the highest stage, Rogers speaks of decreased distance from feelings in the final stages of therapy. There is no contradiction, however. It is partly a matter of how one turns the metaphor: Rogers means by distance *alienation*, whereas we mean *perspective*. The feelings which Rogers and other therapists teach their patients to perceive as feelings were not previously absent but were perceived as properties of the world rather than as feelings of the person. In *The Authoritarian Personality*, this characteristic was labelled "projectivity."

That therapeutic progress can be viewed both as greater distance from ego and closeness to feelings can be accounted for in terms of the psychoanalytic concept of the temporary splitting of the ego which occurs during therapy (Ekstein, 1956). In requiring verbalization of feelings, psychotherapy both requires and encourages some capacity for realistic self-conception.

Sullivan (1953) traces the origin of perspective on oneself to the interpersonal situation, i.e., communication with another person is the basis for learning to see oneself as others

see one. Therapy is an intimate situation between the patient and one other person, with the patient's inner life as subject for discussion. Surely this is an ideal setting in which to acquire the ability to see oneself as another person would.

The average person just does not have this ability, and it is fallacious to use his self-conception, as expressed in interview or personality test, to evaluate the success of therapy. Similarly, a high correlation between one's self-description and one's description of the ideal person is probably a better indication of Philistine self-satisfaction than of the success of therapy. The findings of several investigators that self-reports are inconsistent with or poorly correlated with other measures of adjustment (Beilin, 1957) and of success in therapy (Kelman & Parloff, 1957) are, we believe, clarified by these considerations.

In view of current interest in individuals who withdraw from therapy at the initial stages or who withdraw their children from child guidance clinics before treatment, we would like to record an hypothesis concerning some proportion of those cases. The postulated sequence of events is that an acutely stressful occurrence leads to temporary breaching of the conventional individual's antipsychological defenses, leading him to seek help in psychotherapy. The antipsychological defenses heal over before it is possible to gain access to the underlying difficulty. In such cases, one would expect on follow-up to get bland, "normal," socially approved self-reports on test and interview, regardless of the outcome of the disturbances which led them to register for treatment. By contrast, the self-critical reports of those who stay in therapy are not a fair indication that therapy has not succeeded.

Scott (1958) has ably summarized the definitions of mental health and mental illness utilized in current research. Although the theme of his paper is different from that of the present note, certain conclusions could be justified from his considerations as well as from our own.

It is by no means obvious where one can find either a normal control group to epitomize

the desired outcome of therapy or a neurotic control group to serve as baseline for evaluating therapeutic success. Verbal self-reports, particularly those elicited by objective personality tests, and possibly also those elicited by brief interviews, are not satisfactory data for evaluation of therapy. They reflect too much the conventionality-sophistication dimension, too little other more important aspects of adjustment.

We have no easy solutions to offer. But it seems better to acknowledge the difficulties of the problem than to draw misleading conclusions from spuriously objective criteria of therapy.

REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- BEILIN, H. The prediction of adjustment over a four year interval. *J. clin. Psychol.*, 1957, **13**, 270-274.
- CHRISTIE, R., HAVEL, JOAN, & SEIDENBERG, B. Is the F scale irreversible? *J. abnorm. soc. Psychol.*, 1958, **56**, 143-159.
- COHN, T. S. The relation of the F-scale to intelligence. *J. soc. Psychol.*, 1957, **46**, 207-217.
- EKSTEIN, R. Psychoanalytic techniques. In D. Brower & L. E. Abt (Eds.), *Progress in clinical psychology*. New York: Grune & Stratton, 1956. Pp. 79-97.
- ERIKSON, E. H. Growth and crises of the healthy personality. In M. J. E. Senn (Ed.), *Symposium on the healthy personality*. New York: Macy Found., 1950.
- HOLLINGSHEAD, A. B., & REDLICH, F. C. *Social class and mental illness*. New York: Wiley, 1958.
- KELMAN, H. C., & PARLOFF, M. B. Interrelations among three criteria of improvement in group therapy: Comfort, effectiveness, and self-awareness. *J. abnorm. soc. Psychol.*, 1957, **54**, 281-288.
- REDL, F., & WINEMAN, D. *The aggressive child*. Glencoe, Ill.: Free Press, 1957.
- REMMERS, H. H., & RADLER, D. H. *The American teenager*. Indianapolis, Ind.: Bobbs-Merrill, 1957.
- ROGERS, C. R. A process conception of psychotherapy. *Amer. Psychologist*, 1958, **13**, 142-149.
- ROSE, A. M. Conscious reactions associated with neuropsychiatric breakdown in combat. *Psychiatry*, 1956, **19**, 87-94.
- SANFORD, N. (Ed.) Personality development during the college years. *J. soc. Issues*, 1956, **12**, 1-71.
- SCOTT, W. A. Research definitions of mental health and mental illness. *Psychol. Bull.*, 1958, **55**, 29-45.
- SULLIVAN, H. S. *The interpersonal theory of psychiatry*. New York: Norton, 1953.

Received October 10, 1958. Prior publication.

CRITIQUE AND NOTES

DISPLACEMENT OF PREFERENCE AS A FUNCTION OF AVOIDANCE-AVOIDANCE CONFLICT¹

HAVA BONNÉ GEWIRTZ

University of Maryland

METHOD

AN EARLIER study (Gewirtz, 1959) has demonstrated that approach and avoidance gradients of preference could be obtained from children, following positive or negative reinforcement (success or failure) applied to their response to a training task. The present experiment represents an extension of that study, and attempts to establish a condition under which preference *displacement* may occur.

Since the method employed in this study is generally identical with that of the earlier generalization study (Gewirtz, 1959), it is described here only in brief. Except when otherwise stated, the conditions described below apply equally to both Avoidance and Displacement experiments.

Subjects

Twenty children (8 boys and 12 girls) between the ages of six and eight years served as Ss. They were selected at random from the first and second grades of a private school in the Chicago area. In terms of socioeconomic background, achievement, and intelligence level, this sample was considered similar enough to the one employed in the generalization study to permit relevant comparisons between some of the experimental results obtained from the two studies.

Materials

Five formboard-type puzzles were constructed in such a way that their shapes constituted a similarity dimension, ranging from a square diamond-shaped groove to a long and narrow diamond-shaped one. The two puzzles representing the extremes of the dimension, i.e., Puzzle 1 and Puzzle 5, were used as training puzzles. Each was equipped with a set of seven pieces which could be fitted into the groove. But while this task appeared simple, the asymmetry of the pieces made solution extremely difficult—and actually impossible—for children of the age group employed.

Experimental Treatment

The experimental situation involved a puzzle-solution game played in the presence of the experimenter (*E*). The treatment condition consisted of negative reinforcement applied to *S*'s attempts to solve the assigned training puzzle, when solution was given up by *S* either spontaneously or following *E*'s suggestion (this was considered as one trial). Negative reinforcement represented a combination of (a) actual failure in puzzle solution, (b) mild verbal reproof by *E*, and (c) withdrawal of a small prize (plastic trinket) which had been promised earlier as a reward for successful solution.

General Procedure

Ss were taken individually from their classrooms for the period of the experiment, which lasted approximately 20 minutes. In order to acquaint *S* with the similarity dimension, he was first shown the five puzzles, arranged in a row in random order, and was asked to

Displacement, as the term is used here, describes the phenomenon in which responses shift from the conditioned stimulus towards less similar stimuli, as a function of conflicting tendencies associated with the conditioned stimulus. Miller (1944, 1948) has used the concepts of stimulus generalization (approach and avoidance gradients) in developing his theory of conflict and displacement. While "approach-avoidance" conflict is his major model for displacement, he has also advanced the possibility of displacement resulting from "avoidance-avoidance" conflict. In this case, responses to stimuli at both extremes of a stimulus dimension have been negatively reinforced, producing two avoidance gradients opposite in direction. And when *S* is forced to respond (i.e., when there is no alternative to responding), it is expected that his response would be displaced to a stimulus corresponding to the position where the two avoidance gradients intersect. If these gradients are assumed to be of equal strength at their respective points of origin, the point of their intersection would correspond to the middle of the stimulus dimension. Thus, the resulting displacement curve is expected to have its peak around the central stimulus on the dimension employed.

The present study represents an adaptation of this avoidance-avoidance conflict model to a children's play situation. Its first phase, referred to as the *Avoidance* experiment, is designed to ascertain the shape of the response pattern generated by the negative reinforcement condition employed. In its second phase, referred to as the *Displacement* experiment, the conflict treatment is applied; and it is expected that response displacement will result.

¹ This study was done while the author was at the University of Chicago.

TABLE 1
PREFERENCE PROPORTIONS FOR AVOIDANCE AND
DISPLACEMENT GROUPS

Group	N	Preference Proportions					Tests of Equality ^a	
		Puzzles						
		1	2	3	4	5	χ^2_{4df}	p
Avoidance	10	.02	.04	.06	.20	.68	52.29	<.001
Displacement								
Sequence 1-5	5	.13	.30	.47	.09	.02	—	.0003
Sequence 5-1	5	.10	.32	.40	.13	.05	—	.0083
Combined	10	.12	.31	.44	.10	.03	36.61	<.001

^a When $N = 5$, *p* values were obtained directly from available tables (Bradley & Terry, 1952).

rearrange them in terms of their similarity to each other. The experimental treatment followed this introduction, and is described below separately for the Avoidance and for the Displacement experiments. Following the experimental treatment, *E* presented all five puzzles to *S* by the paired comparison method, and asked him, after each of the 10 pair presentations, to indicate which of the two puzzles in each pair he preferred to play with most. The differential preference rankings assigned to the five puzzles constituted the response measure.²

Avoidance experiment procedure. Ten *Ss* were used in the experiment. Each was given a single negative reinforcement trial on Puzzle 1.

Displacement experiment procedure. Ten *Ss* were used in this experiment. Each was given two negative reinforcement trials, one with each training puzzle (1 and 5). After the first trial, *S* was urged to try and solve the other, "different," puzzle. In order to control for possible sequence effects, five *Ss* had their first trial with Puzzle 1 and five *Ss* had their first trial with Puzzle 5.

RESULTS

The composite response of each treatment group was analyzed by means of the rank analysis method (Bradley & Terry, 1952), which enabled assessment of the differences among the preference rankings assigned to the five puzzles (tests of puzzle preference equality), as well as assessment of differences in response pattern between groups of *Ss* (tests of agreement). Table 1 presents the

² According to the scoring method employed, the preferred member in the pair of puzzles received the rank of 1, and the rejected member received the rank of 2. Each of the five puzzles appeared a total of four times in the 10 pair presentations. Hence, for a single *S*, the sums of ranks for the five puzzles ranged between the values of 4 and 8, i.e., for the most and least preferred puzzle, respectively. Composite (group) scores were based on group averages of such individual sums of ranks.

composite preference proportions (derived from the paired comparison rankings) for the five puzzles, both under the Avoidance and under the Displacement conditions, and the results of the respective tests of puzzle preference equality.

Avoidance Experiment

As seen in the first row of Table 1, the negative reinforcement applied to solution attempts with Puzzle 1 generated a preference pattern which deviated significantly from one where all puzzles are preferred equally. Preference for the puzzles increased with their increasing dissimilarity to the training puzzle, indicating that a steep and consistent avoidance gradient was produced.

For the purpose of the present study, it is also necessary to postulate that a similar avoidance gradient, opposite in direction, would result from a negative reinforcement trial applied to the opposite extreme of the dimension, i.e., to Puzzle 5. Since this assumption appears strongly supported by the findings of the Generalization study,³ it was considered reasonable to assume that the avoidance gradient found here, if reversed in direction, would represent adequately the hypothetical gradient which would be generated by the same treatment when applied to Puzzle 5.

Displacement Experiment

Group data. The data were analyzed first separately for each of the two sequence subgroups (with five *Ss* in each). As shown in Table 1, each subgroup responded to the treatment condition with differential preferences for the five puzzles which deviated significantly from equality. A test of agreement indicated no significant differences $\chi^2 = 1.36$, $p > .80$, 4 *df*) between the two subgroup patterns, showing that order of trials with the two training puzzles did not affect the response. Hence, the subgroups' preference rankings were combined for all further analyses. The preference proportion values obtained from this Combined Displacement group (as well as sepa-

³ It was found that (a) negative reinforcement (three trials) applied to Puzzle 5 generated a reliable avoidance gradient, opposite in direction and similar in slope to the one generated by the identical treatment applied to Puzzle 1; and (b), that the three trials of negative reinforcement applied to Puzzle 1 produced a reliable avoidance gradient which was very similar to the one generated by a single trial of negative reinforcement on the same training puzzle. These comparisons suggest that the negative reinforcement treatment was more effective in determining the response pattern than was the particular training stimulus employed, or even the number of trials. Hence, the assumption above appears reasonable.

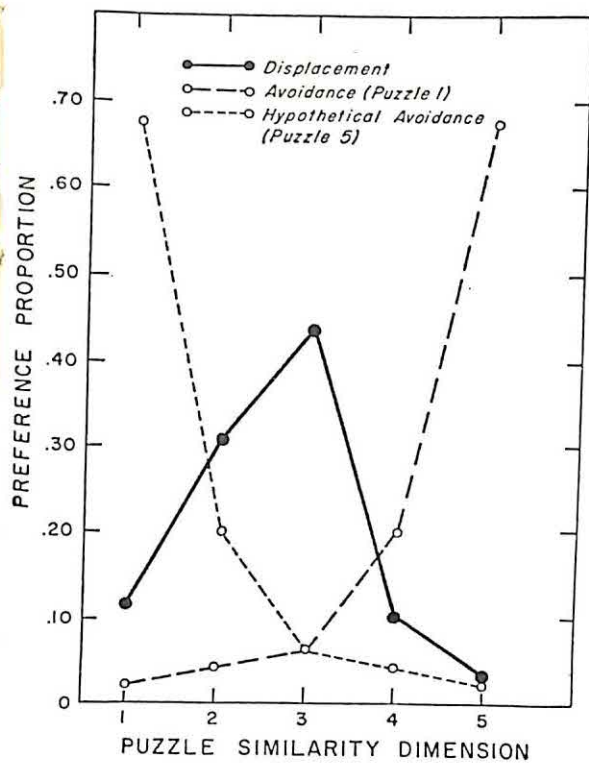


FIG. 1. COMBINED DISPLACEMENT CURVE, AVOIDANCE GRADIENT ORIGINATING AT PUZZLE 1, AND HYPOTHETICAL AVOIDANCE GRADIENT ORIGINATING AT PUZZLE 5 ($N = 10$ PER GROUP)

rately from each subgroup) indicate that Puzzle 3, the central puzzle on the dimension, was the most preferred, while Puzzles 5, 4, and 1 were the least preferred. These differential preferences constituted a displacement pattern which deviated significantly from equality. The Combined Displacement curve, the avoidance gradient generated by negative reinforcement applied to Puzzle 1, and the hypothetical avoidance gradient originating at Puzzle 5 (which is taken to be the mirror image of the Puzzle 1 avoidance gradient) are all illustrated in Fig. 1.

Individual data. More precise evidence for the occurrence of the displacement pattern was sought by an examination of Ss' individual response patterns. For this purpose, S was said to exhibit a "displacement pattern" when the values of his sums of ranks for Puzzles 1 and 5 were larger than those for the three intermediate puzzles (e.g., 8, 6, 4, 5, 7 or 7, 5, 5, 6, 7 for the five puzzles arranged in terms of the similarity dimension; see Footnote 2). A "perfect" individual displacement pattern was defined as one in which maximal preference was given to Puzzle 3 and minimal preference given to Puzzles 1 and 5 (i.e., sums of ranks representing the values 7, 6, 4, 6, 7). Table

TABLE 2
NUMBER OF Ss EXHIBITING DISPLACEMENT PATTERNS WITHIN THE DISPLACEMENT GROUP AND OTHER TREATMENT GROUPS

Group	Pattern		Difference from Displacement Group ^a	
	Displacement	Non-displacement	χ^2_{1df}	p
Displacement	7	3	—	—
Avoidance	0	10	—	.002
Nonreinforcement ^b	5	15	—	.024
Positive ^b	5	35	11.52	<.001
Negative ^b	2	38	18.71	<.001

^a When $N \leq 20$ in the group compared with the Displacement group, exact probabilities were obtained from available tables (Latta, 1953).

^b These three groups are from the generalization study (Gewirtz, 1959).

2 presents the distribution of displacement patterns within the Displacement group, the Avoidance group, and three other groups taken from the Generalization study.⁴ One-tailed tests of independence were employed to compare the proportion of individual displacement patterns found within the Displacement group with that proportion found within each of the other four groups. As shown, the proportion of Ss exhibiting displacement patterns in the Displacement group (7/10) was significantly greater in each case than that proportion produced by a number of different, nonconflict treatments (0/10 in the Avoidance group; 5/20 in the Nonreinforcement group; 5/40 in the Positive group; and 2/40 in the Negative group). Even more dramatic is the fact that of the seven displacement patterns found in the Displacement group, five were *perfect* patterns. No S in any of the other groups exhibited a perfect pattern, with the exception of a single case in the Nonreinforcement group. In other words, 50% of Ss in the Displacement group met the most strictly defined displacement criterion, in contrast to less than 1% of Ss in all other groups combined! This analysis appears to confirm the notion that response displacement occurred as a function of the conflict treatment employed.

⁴ The following groups were taken from the Generalization study: *Nonreinforcement* (Ss were not subjected to experimental treatment and served as a control group); *Positive* (Ss received three trials of positive reinforcement, i.e., successful solution, on the same training puzzle); and *Negative* (Ss received three trials of negative reinforcement on the same training puzzle).

SUMMARY

An experiment on preference displacement was designed along the lines of Miller's avoidance-avoidance conflict model, and adapted to a puzzle-playing situation with children. Ten Ss were negatively reinforced on their attempts to solve the two puzzles representing the extremes of a five-puzzle series which constituted a shape similarity dimension. Ss' differential preferences for the five puzzles were obtained subsequently through paired comparison presentations.

Using a group of 10 different Ss, it was shown that a single negative reinforcement trial applied at one extreme of the puzzle similarity dimension generated an avoidance gradient; and it was assumed that a similar gradient would be produced if an identical treatment were applied at the opposite extreme of the similarity dimension. On this basis, it was expected that the most likely resolution of the conflict between the two incompatible avoidance tendencies resulting from the double negative reinforcement applied in the displacement experiment would be at the point where the two gradients intersect; i.e., response peak was expected around the central stimulus of the similarity dimension. This expectation was supported

by the results, in that a displacement preference curve was obtained: maximal preference was given to the intermediate puzzle, while the puzzles at the extremes of the dimension were the least preferred. It was also found that this treatment condition produced proportionately more individual displacement patterns than did other treatments not involving the conflict condition.

REFERENCES

- BRADLEY, R. A., & TERRY, M. E. Rank analysis of incomplete block designs. I: The method of paired comparisons. *Biometrika*, 1952, **39**, 324-345.
- GEWIRTZ, HAVA B. Generalization of children's preferences as a function of reinforcement and task similarity. *J. abnorm. soc. Psychol.*, 1959, **58**, 111-118.
- LATCHA, R. Tests of significance in a 2 x 2 contingency table: Extension of Finney's table. *Biometrika*, 1953, **40**, 74-88.
- MILLER, N. E. Experimental studies of conflict. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. New York: Ronald, 1944. Pp. 431-465.
- MILLER, N. E. Theory and experiment relating psychoanalytic displacement to stimulus-response generalization. *J. abnorm. soc. Psychol.*, 1948, **43**, 155-178.

Received January 23, 1958.

PERSONALITY, PERCEPTUAL DEFENSE, AND STEREOSCOPIC PERCEPTION¹

JOHN M. DAVIS

School of Medicine, Yale University

WHEN stereograms that present a different picture to each eye are shown through a stereoscope, subjects may report seeing the two pictures in alternation, one superimposed on the other, a mixture of both pictures, or only one of the pictures (in which case the other picture is said to be suppressed). Engel (1955) has suggested that so used, the stereoscope may be a valuable tool for research on relationships between perception and personality and has shown that the meaningful content of the stereoscopic stimuli are important determinants in stereoscopic perception. Cantril (1957) has discussed the stereoscope as an instrument for investigating the role of perception in interpersonal relations. Cross-cultural differences of perception were investigated by Bagby (1957), using American and Mexican subjects, who were presented with stereograms con-

sisting of American scenes paired with Mexican scenes. The scenes from the subject's own culture tended to be perceptually predominant.

This study is an investigation of stereoscopic perception, using an experimental design similar to those employed in tachistoscopic studies of perceptual defense. In the first experiment, some of the variables involved in stereoscopic suppression were investigated, while the second experiment demonstrated differences in the stereoscopic perception of normals and schizophrenics.

EXPERIMENT I

This experiment investigated the effects on stereoscopic perception of the familiarity and emotional content of words, as measured by the Thorndike word count and word association time, respectively. The hypotheses to be tested, which were based on the results of tachistoscopic experiments, were: (a) that words with a low frequency of use would be suppressed more than words with a high frequency of use; and (b) that words with strong emotional content would be suppressed

¹ I would like to express my appreciation for help in preparation of this paper to E. Engel, A. Alderstein, T. Cornsweet, N. V. Bolin, D. Terry, M. Varley, H. Barry, C. Slack, K. Pribram, A. Lemoine, D. E. Hunt, and A. K. Myers.

more often than words with little emotional content.

Method

Stereograms were shown to 30 male volunteer college students through a Keystone Model 40 prism stereoscope. Subjects (Ss) adjusted the slide holder so that a stereogram, consisting of a horizontal line in front of the left eye and a vertical line in front of the right eye, was seen as a cross, giving the proper distance for normal binocular fusion. A card used to cover the stereograms was raised to expose the stimulus word for approximately one second.

Each stereogram consisted of two different typed words, one of which was seen by the left eye and the other by the right eye. Eight stereograms consisted of words of identical word frequencies on the Thorndike (1932) word count, and 16 words were matched with words of different frequencies. Of these 16 pairs, 8 had the more frequent word on the left, while 8 had the more frequent word on the right. The order of presentation of the word pairs was random, as is shown in Table 1. The series of stereograms was presented to one half the Ss as it appears on Table 1; the other half of the Ss were shown the series with the pairs reversed in position (i.e., words appearing on the left on Table 1 were presented to the right eye, and vice versa). After all Ss had seen the 24 stereograms once, the entire series was repeated with the words reversed with respect to the eyes (i.e., all words seen by the right eye in the first presentation were seen by the left in the second showing, and vice versa). The randomization of initial presentation and the systematic reversal serves to control for eye dominance and other variables irrelevant to the present hypotheses.

The Ss were asked always to look with both eyes and to report what they saw, spelling words of which they were not sure. If on both presentations of a given stereogram the S reported seeing the word on the right, for example, the response to that particular stereogram was said to have been the result of right-eye dominance and was not included in the analysis of the data. After the Ss had completed the stereoscopic series, a word association test using the words appearing on the stereograms was administered, and word association time was recorded as a measure of emotional content.

Results

In order to control for the emotional content of stimulus words, the data on pairs of words in which there was at least one word association time

TABLE 1

THE STIMULUS WORDS AND THEIR WORD FREQUENCIES (The word count of Thorndike (1932) gives the rank order of the words by thousands according to their frequency and range.^a The lower the word rank (given parenthetically), the more frequent the word in written language. The pairs of words are numbered in the order of their original presentation, the first word of each pair being that which appeared originally on the left.)

1. friend (1) and French (1)	2. clef (18) and blue (1)
3. bottle (2) and breath (2)	4. such (1) and suck (4)
5. hot (1) and cut (1)	6. cockroach (18) and champagne (13)
7. make (1) and fake (1)	8. kiss (1) and hand (1)
9. love (1) and baby (1)	10. gun (1) and pot (1)
11. articulate (8) and chandelier (13)	12. smear (13) and earth (1)
13. clergy (8) and except (1)	14. vulcanize (18) and brilliant (4)
15. bowel (4) and byway (13)	16. arrive (1) and caster (13)
17. enviable (13) and grandeur (8)	18. daughter (1) and mobilize (18)
19. calcium (8) and viscous (18)	20. fad (13) and lay (1)
21. dance (1) and dream (1)	22. cleave (4) and climax (8)
23. dope (13) and ergo (18)	24. defer (8) and bible (4)

^a When this experiment was being carried out, only the 1932 edition of Thorndike's *Word Book* was available. The word frequencies given here have since been compared to those of the *Large Magazine Count* given in the 1944 edition of the *Word Book*; reanalysis of the data produced no change in the significance levels of the results.

of 2.5 seconds or over were analyzed separately from the data on those pairs of words in which the word association time of both words was under 2.5 seconds (see Table 2). When the responses to stereograms containing words of unequal frequency on which both words had association times under 2.5 seconds were analyzed, it was found that 10 Ss suppressed² more words with the lower frequency of usage according to the Thorndike word count, while only 2 Ss suppressed more words with higher frequency of usage. Three Ss suppressed equal numbers of words with lower or higher frequencies. By sign test (Siegel, 1957), this result is significant at the .038 level (two-tailed). Thus when word association time is below 2.5 seconds, words of low frequency are suppressed.

On the stereograms on which at least one of the words had an association time of 2.5 seconds or over (i.e., on which the S blocked), 5 Ss suppressed more words with the lower frequency of usage, 3 Ss expressed more of such words, and 4 Ss ex-

² A *suppressed* word is defined here as one which was either inaccurately reported or not reported at all, even though it was exposed before each eye. An *expressed* word refers to a word that was reported correctly in at least one of the viewing conditions. When a word is suppressed, the expressed word, a meaningless mixture of letters, or another word not on the stereogram is reported.

TABLE 2

THE RELATION OF WORD FREQUENCY TO SUPPRESSION

Subject	Stereograms Containing No Words on which S Blocked ^a			Stereograms Containing at Least One Word on Which S Blocked ^a		
	No. of times more frequent word expressed	No. of times more frequent word suppressed	Sign	No. of times more frequent word expressed	No. of times more frequent word suppressed	Sign
#1	5	0	+	4	2	+
#2	2	0	+	1	1	
#3	1	0	+	1	1	
#4	0	1	-	0	0	
#5	0	1	-	2	1	+
#6	3	2	+	1	0	+
#7	2	1	+	0	1	-
#8	5	1	+			
#9	4	3	+	0	1	-
#10	1	0	+	0	1	-
#11	1	1				
#12	1	1				
#13	1	1		1	1	
#14	4	0	+	1	0	+
#15	1	0	+			
#16				1	0	+
#17				1	1	

Note.—This table shows the total number of responses of all Ss after eye dominant responses have been deleted. Where no entry appears for an S, all of his responses in that category were eye dominant by this definition.

^a "Blocked" refers to word association times of over 2.5 seconds.

pressed and suppressed equal numbers of the less frequent words (see Table 2) ($p = .73$, two-tailed sign test). The relationship between word frequency and suppression thus does not hold for word pairs in which at least one member has strong emotional content for the S, as indicated by long association time.

To see whether a relationship existed between word association time and word frequency, words with association times of 2.5 seconds or over were divided into two groups according to frequency, the first including words of Thorndike rank from 1,000 to 4,000, and the second those ranked from 8,000 to 18,000. It was found that four Ss blocked more often on words in the category of higher frequency, while six Ss blocked more often on the less frequent words. A sign test performed on these two groups yields a p of .75 (two-tailed). This experiment does not, therefore, show a relationship between word frequency and word association time by this criterion.

It was found, however, that word association time was related to suppression on a stereogram. Nine Ss suppressed the majority of words on which they blocked in the word association test, three Ss suppressed and expressed equal numbers of these

TABLE 3

THE RELATION OF WORD ASSOCIATION TIME TO SUPPRESSION

Subject	No. of Times Words with Association Times of 2.5 Seconds or Over were Expressed	No. of Times Words with Association Times of 2.5 Seconds or Over were Suppressed	Sign
#1	1	5	-
#2	1	1	
#3	1	1	
#4	0	1	-
#5	0	3	-
#6	0	1	-
#7	1	0	+
#9	1	1	
#10	0	1	-
#13	0	2	-
#14	0	1	-
#16	0	1	-
#17	0	2	-

Note.—The data given here constitute all of the responses to stereograms containing words on which the S blocked, except those classified as eye dominant responses.

words, and only one S expressed the greater number of them (See Table 3). This difference is significant at the .022 level (two-tailed sign test).

It was not possible to randomize the positions of the words with long association times, because they were not determined until after the word pairs had been presented. Therefore a study was made of the patterns of suppression responses. All Ss clearly showed eye dominance in the majority of their responses. Of the Ss who suppressed one or more words with emotional content, 6 more often suppressed words which appeared first on the eye dominant side, while 5 more often suppressed words first appearing on the nondominant side.

EXPERIMENT II

Method

Six stereograms were shown to 5 male and 5 female psychotics, patients in a private mental hospital, and to 10 normal Ss, 6 male and 4 female, medical patients. Both groups of Ss were in the lower middle class between 20 and 50 years of age; however, the normal group averaged about 10 years younger than the psychotic group.

Each stereogram consisted of two different pictures cut from popular magazines, arranged in such a way that the meaningful parts of each picture stimulated roughly corresponding parts of the two retinæ. The series of stereograms was shown to the Ss twice; in the second presentation the pictures were reversed with respect to the eyes as a control for eye dominance. The stereograms were presented to the S for a period of one minute. The Ss were asked to describe what they

TABLE 4
DESCRIPTORS OF THE STEREOGRAMS IN
EXPERIMENT II

Picture A	Picture B
1. A young woman's face	A small figurine
2. A man's face	A snake
3. A bowl of Rice Crispies	A Negro man kissing a white girl
4. A man's head	One man stabbing another with a knife
5. A man's head	Two women and a man kneeling in prayer
6. A girl and man in a garden	Two girls and a man in a bathing suit

saw, and their responses were recorded verbatim. Table 4 gives short descriptions of the stereograms.

Responses were classified in five categories. Category 1 (superimposition) consisted of those responses in which the *S* reported seeing both pictures either overlapping or superimposed on each other. Category 2 (alternation) consisted of responses in which the *S* reported seeing the pictures in alternation. Sometimes the pictures would alternate rapidly, and sometimes one picture would slowly fade to be replaced by the other. In Category 3 (suppression) were classified responses in which the *S* saw the same picture on both presentations. In Category 4 (admixture) were placed responses in which he saw an unusual composite of both pictures or something not presented in either picture. The fifth category consisted of eye dominant responses, and was not included in the analysis of differences between psychotics and normals.

Results

On the basis of an unpublished preliminary investigation using psychotics and normals (college students) in a similar testing situation, it was predicted that normals would make more responses in Categories 1 and 3, and that psychotics would give more responses in Categories 2 and 4. When one-tailed Mann-Whitney tests were performed, it was found that psychotics did score significantly higher on Category 4. Normals scored significantly higher on Category 1 and almost significantly higher on Category 3 (see Table 5).

Under the conditions of this experiment, it was clearly evident that different *Ss* reported seeing different things even though the stimuli were identical. For example, many *Ss* reported seeing a bowl of Rice Crispies and a woman on Stereogram 3, and thus did not report seeing the socially taboo Negro man who was kissing the woman. Three psychotics and no normals reported seeing a baseball game on Stereogram 4. It seems likely that they did actually see the knife fight but in-

TABLE 5
COMPARISON OF NORMAL AND SCHIZOPHRENIC SUBJECTS
ON FOUR RESPONSE CATEGORIES

Category	Schizophrenics		Normals		<i>p</i> (one-tailed)
	Mean Frequency of <i>Ss</i> ' Responses	<i>R_s</i>	Mean Frequency of <i>Ss</i> ' Responses	<i>R_n</i>	
1	2.1	78	5.4	132	.025
2	4.0	101.5	2.2	108.5	0
3	1.3	84	3.6	126	.06
4	2.3	131	.6	79	.03

terpreted it as a baseball game; that is, they were sensitive to the aggression but denied the aggressive content of the picture.

DISCUSSION

Gating mechanisms have been shown to be important in some areas of neural function. For example, Galambos (1956) has shown that auditory nerve discharge is reduced or abolished by electrical stimulation of the floor of the medulla. Hernandez-Peon, Scherrer, and Jouvett (1956) have further observed in cats that a reduction in the auditory response of the dorsal cochlear nucleus occurs with the presentation of a distracting visual stimulus (a mouse), an olfactory stimulus (fish odor), or an electric shock to the forepaw. There is evidence that such neurophysiological gating effects may also be present in vision, olfaction, and cutaneous sensation.

The body of evidence indicating the neurophysiological importance of gating mechanisms would suggest that stereoscopic suppression as well as tachistoscopic perceptual defense may be conceptualized in such terms. The gating mechanism employed in these perceptual operations may be similar to that involved in denial and the other defense mechanisms. In the stereoscopic situation the *S* makes a choice between two stimuli, a choice that is presumably unconscious since he is not aware of the gated-out stimulus. The stereoscopic technique may thus provide an experimental instrument of some importance for the investigation of psychological defense.

SUMMARY

The present paper reports on two experiments on stereoscopic perception. Thirty subjects were shown 24 stereograms, each of which presented a different word to each eye. Association times for the words presented were measured for all subjects as a measure of emotional content, and Thorndike word count ratings were available. Words with low frequency of usage and words with long association time tended to be suppressed. In the

second experiment, some observations made on psychotics and normals indicated that the stereoscopic perception of psychotics differed significantly from that of normal subjects. It is suggested on the basis of this preliminary evidence that the stereoscopic technique might prove experimentally useful, particularly in the study of psychological defense.

REFERENCES

- BAGBY, J. W. A cross-cultural study of perceptual predominance in binocular rivalry. *J. abnorm. soc. Psychol.*, 1957, **54**, 331-334.
- CANTREL, H. Perception and interpersonal relations. *Amer. J. Psychiat.*, 1957, **114**, 119-126.
- DAVIS, J. M. Four studies in personality. *Education Testing Service*, RM-57-6, 1957.
- ENGEL, E. Meaningful content in the study of rivalry

and fusion—preliminary observations. Unpublished manuscript, Dept. of Psychology, Princeton University, 1955.

- GALAMBOS, R. Suppression of auditory nerve activity by stimulation of efferent fibers to cochlea. *J. Neurophysiol.*, 1956, **19**, 424-437.
- HERNANDEZ-PEON, R., SCHERRER, H., & JOUVET, M. Modification of electrical activity in cochlear nucleus during "attention" in unanesthetized cats. *Science*, 1956, **123**, 331-332.
- SIEGEL, S. *Nonparametric statistics for the behavioral sciences*. New York: McGraw-Hill, 1956.
- THORNDIKE, E. L. *A Teacher's word book of 20,000 words*. New York: Columbia Univer. Press, 1932.
- THORNDIKE, E. L., & LORGE, I. *The teacher's word book of 30,000 words*. New York: Columbia Univer. Press, 1944.

Received February 6, 1958.

THE RELATIONSHIP BETWEEN GSR AND SPEECH DISTURBANCES IN PSYCHOTHERAPY

DAVID M. PANEK AND BARCLAY MARTIN

University of Wisconsin

THE development of procedures for obtaining continuous measures of momentary fluctuations in anxiety level is important for research in psychotherapy. Two such methods have been employed recently. Mahl (1955, 1956a, 1956b) quantified expressive aspects of patient speech, including such categories as "Ahs," "Repetitions," "Sentence Corrections," "Stutters," etc., and found in an analysis of one case that these speech disturbances increased when the patient talked about content which had been independently assessed as anxiety arousing. Dittes (1957a, 1957b) recently employed the GSR as a measure of anxiety and found in a study based on one patient that GSRs accompanying embarrassing sex statements tended to extinguish as psychotherapy progressed. In another study, also based on one patient, he found that GSR activity increased when the therapist became less permissive. If both speech disturbances and the GSR reflect momentary changes in anxiety level, or somewhat more broadly speaking, emotional arousal, one would expect the two measures to covary in a systematic manner, and such is the general hypothesis of the present study.

This research was also concerned with investigating more specific aspects of the hypothesized relationship between the GSR and speech disturbances. For example, it was thought that the category of Repetitions, defined in terms of the unnecessary repetition of words or short phrases,

may reflect the immediate disrupting effects of anxiety, whereas Ahs may reflect a more controlled and effective defense against the disrupting effects of anxiety. If such were the case, one would expect Repetitions to be maximal at the point of greatest emotional arousal as indicated by the GSR. Ahs, on the other hand, may indicate the person's attempt to recover from the disrupting effects of anxiety and would accordingly be maximal at a point displaced in time from the point of maximum emotional arousal. This possibility was further suggested by Mahl's (1955) empirical finding that the frequency of occurrence of Ahs tended to be negatively correlated with the occurrence of Repetitions among subjects.

METHOD

Subjects. Four interviews from each of four psychotherapy clients seen by psychologists at the University of Wisconsin provided the basic data for this study. The four clients (three males and one female) were between 19 and 31 years of age. Two had recently begun psychotherapy; two had been in psychotherapy for over a year. Synchronized GSR and tape recordings had been routinely obtained for these clients for most of their therapy sessions. For purposes of this research, four interviews were selected for each client from those for which there were technically satisfactory GSR and tape recordings. The criterion for this selection was the presence of a

relatively large number of scorable GSR dips as defined below. In all parts of the research, analysis of the GSR data was kept independent of the speech disturbance analysis.

Instrumentation. The GSR apparatus was an AC circuit designed by Grant (1946), which used dry silver electrodes attached to the palmar surfaces of two fingers and fed into a G.E. Photoelectric Recorder. At the beginning of therapy, the therapist tapped a telegraph key three times, which simultaneously introduced audible clicks on the tape recorder and made a mark on the GSR record. At the end of the session, the therapist again tapped the telegraph key to provide a final synchronizing signal. There were no apparatus manipulations to concern the therapist between these signals at the beginning and end of the session.

The speech disturbance measures. The most frequently occurring speech disturbances in Mahl's work were Ahs, Repetitions, and Sentence Corrections. In the present study, the speech disturbance scores were obtained directly from the tape recordings rather than from typed transcripts, and preliminary work indicated that Sentence Corrections could not be reliably scored in this way. Accordingly, scores were obtained for only Ahs and Repetitions. Ahs were scored whenever the definite "ah" sound occurred. Repetitions were scored when there was a serial repetition of one or more syllables or one to two words.

The 16 interviews were scored for the number of Ahs and Repetitions occurring in each 30-second interval beginning at the initial synchronizing signal. This was accomplished by connecting a footswitch via a Hunter Timer to the Revere tape recorder so that after a depression of the foot pedal, the tape recorder played for exactly 30 seconds and stopped. The interviews were then played through a second time, and the time that the client and the therapist spent talking in each 30-second interval was obtained by using two time clocks. The speech disturbance scores were divided by the time the client spent talking to yield a rate of speech disturbance as the final score used in statistical analyses.

The GSR measures. The selection of the specific GSR dips to be used in this analysis was made after all interviews had been scored for speech disturbances. The GSR records were marked off in 30-second intervals corresponding to the same intervals used for the speech disturbance scoring. The records were then examined for distinct "dips," that is, those dips relatively isolated in time from preceding or subsequent dips. For purposes of investigating the relationship between GSR and speech disturbance, it was decided to divide the intervals around the GSR dip in the

following manner: The 30-second interval in which the dip happened to fall was called the 0 interval, the two 30-second intervals preceding the 0 interval were classified as -1 and -2, and the two intervals following were classified as +1 and +2.

It was not possible to obtain a sufficient number of distinct and isolated GSR dips, and some compromises were made. The following rather arbitrary criteria were used in selecting usable GSR dips:

1. The dip must have a magnitude of at least 2200 ohms.
2. A 30-second interval was not scorable for speech disturbances if it occurred within two intervals of a succeeding or preceding dip other than the dip being studied.
3. There must be at least three scorable 30-second intervals including the 0 interval associated with a dip. This meant that for some dips, speech disturbance scores were not obtained for one or two of the five possible intervals.
4. A dip was used when other dips were within the surrounding intervals if the selected dip was at least twice as large as those surrounding it and if the dips tended to increase in size approaching the 0 interval, suggesting a high point of emotional arousal.
5. All intervals in which the client spent less than five seconds talking were considered unscorable because the rates of speech disturbances would be based on too small a sample of client talk.

The relationship between the GSR and speech disturbances was further examined by selecting from each interview on the basis of the GSR record the five continuous minutes of maximal GSR activity and the five continuous minutes of minimal GSR activity. One interview was not usable because the GSR activity was consistently high throughout. Thus, the speech disturbance scores used in this comparison were based on the average score for the four interviews for three clients and three interviews for one client.

RESULTS

All speech disturbances were scored by one judge. The reliability of scoring was assessed by having a second judge independently score the first 20 minutes of two interviews. The correlations between the two judges' scores on the 40 30-second intervals involved in each of the two interviews were .87 and .88 for Ahs and .65 and .78 for Repetitions. As an additional reliability check, a third interview was carefully transcribed after the two judges had scored it, with great care being taken to get all Ahs and Repetitions in the typed transcript. The correlations of the two judges' scores with scores obtained from the typed manuscript were .98 and .90 respectively for Ahs and .94 and .76 respectively for Repetitions. It

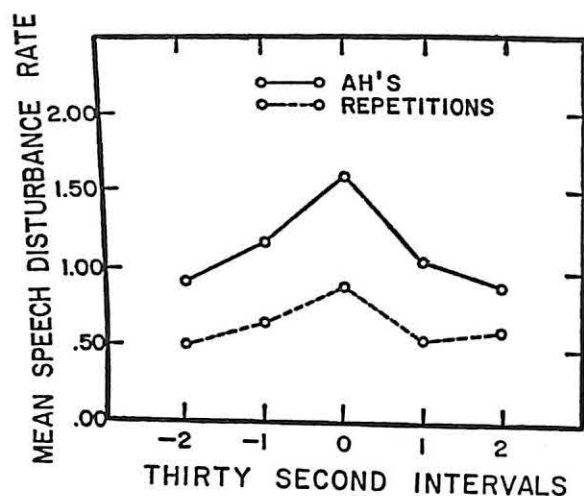


FIG. 1. SPEECH DISTURBANCE RATE AS A FUNCTION OF INTERVALS AROUND GSR DIP

was concluded that the procedure of scoring Ahs and Repetitions directly from the tapes had adequate reliability with the Ahs being somewhat more reliably scored than Repetitions.

Figure 1 shows the mean rate of Ahs and Repetitions as a function of the 30-second intervals surrounding the GSR dip. It is apparent that both Ahs and Repetitions reached a maximum during the 0 or dip interval. An extension of Alexander's test for trend as developed by Grant (1956) was employed to see if the trends manifested in these curves were significantly different from chance. Summary tables for the analyses of variance with respect to Ahs and Repetitions are shown in Tables 1 and 2. The dependent variable used in these analyses was always the average score for a client based on the data for all four interviews, which then yielded four scores representing the four clients at each of the five intervals. For Ahs there is a significant quadratic component, which indicates that the elevation of the curve in the center compared to the more extreme intervals is probably not due to chance. For the Repetitions there are significant quadratic and quartic components, which indicates that it is unlikely that the central elevation as well as the flattening at the extreme intervals is a chance finding. There is no evidence in support of the notion that Ahs and Repetitions would be maximal at different points in time with respect to the GSR dip; both measures reach a maximum at the 0 interval.

One result of scoring the speech disturbances in the intervals around the GSR dip by the present procedure was that the average speech disturbance

TABLE 1
SUMMARY OF TREND TESTS FOR MEAN RATE OF AHS

Source	df	Mean square	F
Over-all			
Slope	1	.010	0.7
Quadratic	1	.886	16.1*
Cubic	1	.020	0.9
Quartic	1	.400	4.4
Between Individual			
Slope	3	.013	—
Quadratic	3	.055	—
Cubic	3	.023	—
Quartic	3	.090	—

* Significant at .05 level.

TABLE 2
SUMMARY OF TREND TESTS FOR MEAN RATE OF REPETITIONS

Source	df	Mean square	F
Over-all			
Slope	1	.006	0.7
Quadratic	1	.155	11.1*
Cubic	1	.050	1.2
Quartic	1	.156	12.0*
Between Individual			
Slope	3	.884	—
Quadratic	3	.014	—
Cubic	3	.043	—
Quartic	3	.013	—

* Significant at .05 level.

scores were based on different numbers of GSR dips at the five different intervals. Thus, for the intervals running -2, -1, 0, +1, +2, the total number of dips for which there were scorable intervals at these five points were 39, 69, 84, 74, and 42, respectively. These unequal *N*s might result in greater variability in those intervals with the fewer cases but presumably would not affect the means in any constant way. As a check on this possibility, however, another analysis was performed, using the three middle intervals only, with the necessary GSR dips omitted so as to equalize the number of dips involved. This resulted in the scores for the three intervals being based on 59 dips. The quadratic component was still significant for a combined speech disturbance score based on both Ahs and Repetitions, but it did not reach significance for the separate analyses of Ahs and Repetitions. However, the reduction in the number of dips used undoubtedly increased the variability, and also degrees of freedom were lost in the analysis as a result of eliminating the two extreme intervals.

There were no significant differences between mean speech disturbance scores obtained in the five-minute segments of highest GSR activity as compared with those obtained in five-minute segments of lowest GSR activity, although the differences were in the expected direction ($t = 1.24$ for the combined speech disturbance score). Although the differences were small, in 13 of the 15 interviews the mean rate of combined speech disturbances was higher in the high GSR activity segments than in the low GSR segments.

DISCUSSION

The basic hypothesis of this research that speech disturbances would show an increase at points of emotion arousal as defined by the GSR was confirmed, thus providing some integration of the work of Mahl (1955, 1956a, 1956b) on speech disturbances and Dittes (1957a, 1957b) on GSRs during psychotherapy. The more gross comparison of five-minute segments of high and low GSR activity did not yield significantly different speech disturbance scores. There was also no evidence that Ahs and Repetitions were differently related to points of emotional arousal.

A more general implication of these findings is with respect to the significance of the individual GSR dips. Earlier studies (Bixenstine, 1955; Martin & McGowan, 1955; Mowrer, Light, Luria, & Zeleny, 1953) have frequently focused upon changes in general level of skin conductance or sweating as an indicator of anxiety level or emotional arousal. The fact that in this study discrete GSR dips were found to be related to increases in speech disturbances suggests that in psychotherapy sessions these discrete dips do not, by and large, result from extraneous external and internal stimuli that are irrelevant to the process of therapy. Further analyses of these individual GSR dips in relation to other aspects of the therapeutic process would appear to be indicated. An index based upon both speech disturbance measures and occurrence of GSR dips should prove to be a fairly reliable and valid indicator of the momentary changes in anxiety level in psychotherapy interviews.

SUMMARY

The purpose of this study was to investigate the relationship between points of emotional arousal or anxiety as defined by the GSR and the occurrence of speech disturbances in psychotherapy. Synchronized GSR and verbal recordings of four interviews from each of four psychotherapy clients provided the basic data. It was found that speech disturbances were maximal at times coincident with GSR deflections and decreased in frequency on either side of GSR deflections. This research provides additional validity for both speech disturbances and the GSR as measures of momentary changes in anxiety level in psychotherapy.

REFERENCES

- BIXENSTINE, V. E. A case study of the use of palmar sweating as a measure of psychological tension. *J. abnorm. soc. Psychol.*, 1955, **50**, 138-143.
- DITTES, J. E. Extinction during psychotherapy of GSR accompanying "embarrassing" statements. *J. abnorm. soc. Psychol.*, 1957, **54**, 187-191. (a)
- DITTES, J. E. Galvanic skin response as a measure of patient's reaction to therapist's permissiveness. *J. abnorm. soc. Psychol.*, 1957, **55**, 295-303. (b)
- GRANT, D. A. A convenient alternating current circuit for measuring GSR's. *Amer. J. Psychol.*, 1946, **59**, 149-151.
- GRANT, D. A. Analysis-of-variance tests in the analysis and comparison of curves. *Psychol. Bull.*, 1956, **53**, 141-154.
- MAHL, G. F. The use of "ah" in spontaneous speech. Paper read at Eastern Psychological Association, 1955.
- MAHL, G. F. Disturbances and silences in the patient's speech in psychotherapy. *J. abnorm. soc. Psychol.*, 1956, **53**, 1-15. (a)
- MAHL, G. F. "Normal" disturbances in spontaneous speech: General quantitative aspects. Paper read at American Psychological Association, Chicago, September, 1956. (b)
- MARTIN, B., & MCGOWAN, B. Some evidence on the validity of the Sarason Test Anxiety Scale. *J. consult. Psychol.*, 1955, **19**, 468.
- MOWRER, O. H., LIGHT, B. H., LURIA, Z., & ZELENY, M. P. Tension changes during psychotherapy with special reference to resistance. In O. H. Mowrer (Ed.), *Psychotherapy: Theory and research*. New York: Ronald, 1953. Pp. 546-640.

Received February 19, 1958.

SOME FACTORS IN THE SELECTION OF LEADERS BY MEMBERS OF SMALL GROUPS

JOHN P. KIRSCHT, THOMAS M. LODAHL, AND MASON HAIRE

University of California, Berkeley

SEVERAL studies have reported the relationship between leadership and the rate of participation. Bass (1949) found a correlation of .93 between ratings on leadership and the amount of participation time in ten-person groups. However, the size of this correlation may be partially the effect of using the subjects as raters and of the prior acquaintance of the subjects in a class. A study by Borgatta and Bales (1956) reports that high ratings on leadership by group members tends to be associated with high rates of interaction-initiation. Moderate but consistently positive correlations between amount of participation and sociometric scores were found by Peterman (1950).

Although such findings are not unexpected because of the generally task-oriented behavior of laboratory discussion groups, participation, by itself, may not be very predictive of leadership. Slater (1955) reports that leadership as rated by other group members coincided with highest participation in only 11 of 20 discussion groups.

Other researchers have attempted to specify the particular kinds of interaction that differentiate the leaders of small groups. Carter, Haythorn, Shriver, and Lanzetta (1951), using four-person groups, examined the interaction in 53 scoring categories. Two of these categories consistently differentiated leaders in three kinds of tasks: "diagnoses situation—makes interpretation" and "gives information on carrying out action." Other categories that showed differences but not as reliably were: "proposes course of action for others", "initiates action toward problem-solving which is continued," and "integrates group behavior." It was also found that in the reasoning task, leaders ask for information or facts significantly more often than nonleaders. Furthermore, Shaw and Gilchrist (1956) found that leader rank and the number of written communications sent were positively related and that the major source of the difference for leaders was communications about organizing the group and giving factual information.

Thus it seems that leaders are persons who organize the group, solicit and integrate contributions, and propose courses of action. A high rate of participation would not necessarily be associated with these behaviors.

The present study was designed to investigate the relationship between amount of participation, frequency of task- and group-oriented interaction,

and the selection of leaders by other group members. The general hypothesis was that amount of participation and organizational-integrative interaction are both associated with leader selection but that each may reflect different aspects of the criterion.

METHOD

Two three-person groups met simultaneously in separate rooms and were given identical instructions by the experimenters. Each group discussed a human relations problem for about 20 minutes, reaching a group consensus in that time. The groups were then told that a second meeting would take place immediately in which one person from each group would act as a representative; the two representatives would discuss a concrete problem related to the previously discussed topic. Each group then selected one of the members to act as representative.

Tape recordings of the meetings were scored for the amount of time each participant talked. Also, each meeting was scored with a system of eight categories based on Bales' (1950) system. In general, the unit of analysis was one simple statement. The three relevant categories and their definitions follow:

D. Gives suggestion (any statement which proposes a course of procedure but is *not* simply an expression of opinion, e.g., "Let's have each person give his solution"—not, "I think teachers should unionize").

E. Asks for suggestion, opinion, fact, (any request which is not for repetition or rhetorical purposes, e.g., "What do you think?" or "Would you agree that...").

F. Sums up, integrates (any statement attempting to organize the points covered in the discussion, or to bring together diverging opinions, or to re-word or make explicit opinions of others, e.g., "We seem to agree that strikes are bad"; "These restrictions would qualify what we said earlier").

These three categories will be referred to as "DEF." They define group-oriented, organizational kinds of interaction and are similar to the categories that differentiated leaders in the Carter et al. (1951) study. It was hypothesized that the interaction scored in these categories would distinguish those members who would be chosen by the group as representatives from those not chosen.

Data were obtained on 26 three-person groups.

Two of the groups were discarded because of incomplete interaction records and two were also discarded because they used random devices (flipping coins, etc.) to select the representative.

RESULTS

For the total sample of 66 Ss, data on time-talked were put into percentage form for each meeting. Those members who were chosen as representatives talked an average of 44.8% of the meeting time; nonrepresentatives talked an average of 27.6% of the time. This difference yields a t of 5.2, significant beyond the .001 level. The average number of DEF interactions per meeting was 12.6 for representatives and 6.1 for nonrepresentatives. However, the variances are heterogeneous in this case, so a median test was used, yielding a chi square of 8.79, significant beyond the .01 level. Thus, both on measures of amount of participation and DEF interaction, those members chosen as representatives were higher.

The time-talked measure correlated with DEF scores $+ .39$, p less than .01. Although a moderate, positive relationship exists between participation and DEF, the two variables do not measure the same thing. When the DEF scores were put into percentage of total DEF per group, the correlation between DEF and time-talked is $+ .44$, not significantly higher than the former correlation.

Since the measure of leadership was a discrete variable, point-biserial correlations were computed to determine the amount of relationship between leader choice and the two interaction measures. It was found that the time-talked measure correlated with leadership $+ .543$; for DEF scores, the point-biserial correlation coefficient was $+ .527$. Both of these correlations are significant beyond the .01 level.

A weighted combination of the two interaction variables was made up using the discriminant function (McNemar, 1955, pp. 210-211). The relative weights were .26 for time-talked and .16 for DEF. This combination yields the best discrimination between the distributions of scores for representatives versus nonrepresentatives. When composite scores were made up using these weights, they correlated with leadership $+ .632$.

If the selection of leader is examined group by group, the predictive power of the interaction measures can be determined and compared. That person who talked the most was chosen as representative in 14 of the 22 groups. Since the chance level of selection is one out of three, this represents a highly significant ($\chi^2 = 9.15$, $p < .01$) prediction, if not too reliable in a practical sense. In those groups where one person dominated the discussion, i.e., talked over 50% of the time ($N = 8$), time-

talked accounts for the representative in seven of the eight cases.

In 14 of the 22 cases, the person with the highest DEF score was chosen as representative. In two of the negative cases, two members were tied in DEF score, and one of the two was chosen. There are 10 cases in which the highest DEF score does not coincide with the highest time-talked. Thus, it is apparent that the two variables do not account for the same thing.

If the dominant person is predicted in those cases in which one person talks over 50% of the time, and where this is not applicable, DEF scores are used, 16 of the 22 cases are accounted for. This method of combining the two variables yields the best practical prediction that can be made from the data.

Although one of the DEF categories may be more crucial than the others, none of the three was found to predict as well by itself as the combined score. It may be that there are particular kinds of groups or situations in which different types of integrative behaviors are particularly appropriate, but these refinements are not possible with the present data.

The other categories used in scoring the interaction yielded no significant differences between representatives and nonrepresentatives. These categories included agreement, positive social interaction, tension release, disagreement, and hostility.

DISCUSSION

The major problem in this experiment is the isolation of variables that can be used to account for the perception of leadership by group members. The data show that amount of participation and DEF interaction are significantly related to leadership choice.

Participation and DEF interaction are neither independent nor highly correlated. They show a moderate, positive relationship. Conceptually, we might think of the group discussion situation in functional terms. The task set by the experimenter defines the goal of the group. Analytically, the process of reaching this goal can be broken down into several components, as is done by Bales (1950). One set of problems involves the production of relevant ideas by the group members; a second set is concerned with the organization and integration of these ideas into a solution that is acceptable to the group. If it can be assumed that the perceptions of the group members of each other with respect to leadership center around the group's functioning, then it appears that the relationship between these two aspects of problem-solving is not a simple one. Within the area of perceived leadership are two subregions: amount of participation and organiza-

tional behavior. In general, these two overlap partially, but it appears from our data that the relative importance and the amount of overlap varies from group to group. In some groups, it appears that any person who takes over the task and produces something will be chosen as leader. In other groups, ideas may be plentiful and the problem is to work out a mutually acceptable result. While this scheme makes difficult a simple prediction of leader choice, it appears to do more justice to the data than a monolithic concept of leadership in small groups.

Another way to state this interpretation is in terms of the differentiation of roles in small groups. Where one member talks a great deal, he is perceived as an appropriate representative, but in a more equal participation situation, group-centered, integrative behavior is relatively more important. In the latter case, persons are differentiated with respect to functioning in synthesizing the group. Even in this case, however, there is undoubtedly a minimum amount of participation necessary before the group-oriented behavior is noticed.

It must be remembered that these groups met only once and that the members were unacquainted with each other prior to the meeting. The representatives are leaders only in the narrow sense that they are empowered to act for the group. While these results may not pertain to a wide range of leadership situations, many short term, problem-solving groups are found that are similar to the experimental situation. Furthermore, the results with respect to DEF interaction seem to corroborate the findings of Carter et al. (1951) regarding the types of interaction which are important in leader behavior. This confirmation includes Carter's negative findings also: the expression of

agreement, "personal feelings," and disagreement did not reliably differentiate emergent leaders from nonleaders. In addition, the results on amount of participation tend to confirm the importance of this variable, as has been found by other experimenters.

In summary, the general picture of a small group leader which appears to be common to various researches shows a group member who tends to have a high rate of participation in the discussion; he is task-oriented, attempts to specify the problem, to suggest courses of action, to seek out the members' contributions, to integrate these and to propose solutions in the attempt to secure consensus in the group.

REFERENCES

- BALES, R. F. *Interaction process analysis*. Cambridge: Addison-Wesley, 1950.
- BASS, B. M. An analysis of the leaderless group discussion. *J. appl. Psychol.*, 1949, **33**, 527-533.
- BORGATTA, E. F., & BALES, R. F. Sociometric status patterns and characteristics of interaction. *J. soc. Psychol.*, 1956, **43**, 289-297.
- CARTER, L., HAYTHORN, W., SHRIVER, B., & LANZETTA, J. The behavior of leaders and other group members. *J. abnorm. soc. Psychol.*, 1951, **46**, 589-595.
- MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1955.
- PETERMAN, J. N. Verbal participation, its relation to decision satisfaction and the leader function in decision making groups. Ann Arbor: Univer. of Michigan Conference Research Project, 1950.
- SHAW, M. E., & GILCHRIST, J. C. Intra-group communication and leader choice. *J. soc. Psychol.*, 1956, **43**, 133-138.
- SLATER, P. E. Role differentiation in small groups. *Amer. Sociol. Rev.*, 1955, **20**, 300-310.

Received December 27, 1957.

YIELDING, AUTHORITARIANISM, AND AUTHORITARIAN IDEOLOGY REGARDING GROUPS¹

EUGENE B. NADLER

Case Institute of Technology

THIS investigation was prompted by Asch's (1952) discussion of the personal and social significance of yielding to group pressure. For the person, yielding requires the inhibition of overt and perhaps implicit responses to objective situational

requirements. For the group, malignant processes may arise and develop, unchecked by those members in whom goals of personal security predominate over group goals. In the present study, the view was taken that both meanings of yielding are assigned positive value in two separate ideologies concerning the group-individual relation. One of these is termed *conformity*, which includes the value of submission of the individual to the demands of the group, even if so doing requires the inhibition of response to the objective situa-

¹ This article is based in part on a doctoral dissertation submitted to the Department of Psychology, Western Reserve University, February, 1957. The author wishes to thank George A. Albee for his valuable aid and guidance as chairman of the doctoral committee.

tion. The other is *rugged individualism*. Among other features, this ideology countenances the decay of group structures which may result from the valued expression of egocentric motives. Asch's discussion led to the hypothesis that the extent to which individuals yield is related to the degree of their acceptance of either of these two ideologies.

Previous research also suggested the more inclusive hypothesis that yielding, as well as conforming and rugged individualistic ideology, could be viewed within the conceptual framework of the authoritarian personality. Frenkel-Brunswik (Adorno, Frenkel-Brunswik, Levinson, & Sanford, 1950, Ch. 11) reports that in interviews, people scoring in the highest quartile of the E scale showed acceptance of the idea of survival of the fittest, hero worship of acquaintances, a view of the world as a jungle, and exploitive manipulative opportunism, i.e., rugged individualism. The same people showed conventionalism and the desire for status-acceptable qualities in their friends, i.e., conformity. However, her account does not reveal the extent of relationship between the two patterns, nor how they influence behavior in a concrete situation.

METHOD

Group pressure was simulated by announcing bogus group judgments to small groups of Ss whose apparent task was to match one of three lines of variable length with a standard line (Asch, 1952, Ch. 14). This method was proposed by T. Rice,² and is similar to that independently announced by Crutchfield (1955) shortly after the present study was begun. A more complete description of the method, and the precautions taken to maintain the deception, may be found in Nadler (1957).

Questionnaires were used to measure the other variables, which were: the ideology of conformity (IC), the ideology of rugged individualism (RI), ethnocentrism (E), and authoritarianism (F). The California E and F scales (Adorno et al., 1950) were used to measure the latter two variables. Similar Likert type scales were used to measure conformity and rugged individualism. Reliability and internal consistency of IC and RI had been determined beforehand in 107 college students.

Items in the IC scale stressed the value of compliance with the demands of common reference groups and the necessity for submerging the self in the group. Punishment for dissent was both implied and explicated, as was the hopelessness of dissent. Items in the RI scale denied the importance of

² Personal communication from Western Reserve University graduate student.

TABLE 1
STATISTICAL PROPERTIES OF THE MEASURES

Property	Measure				
	IC	RI	E	F	Yielding ^a
Reliability	.76	.62	.85	.74	.52
Mean	3.60	4.21	3.28	3.70	2.84
SD	.86	.79	.99	.88	2.52
N	91	91	71	91	70
No. of items	20	20	14	14	12

^a Mean and SD of this measure refer to the full scale, each item scored pass-fail; means and SDs of the other measures refer to each item, scored 7-1 on an agree-disagree continuum.

TABLE 2
INTERCORRELATIONS AMONG THE MEASURES

Measure	RI	E	F	Yielding
IC	.40	.55	.68	.30*
RI		.56	.63	.29*
E			.61	.30*
F				.48

* Significant beyond the .025 level. All others significant beyond the .01 level.

social forces, asserting both the undesirability and the impossibility of groups having any decisive influence upon individuals. Other items raised narrow self-interest to the level of a virtue and denied social motives.

Since the original F scale contained many items with conforming or rugged individualist content, it was necessary to omit such items from the version used here, particularly since some of the items had been used in the IC and RI scales.³

The questionnaires were administered to 91 college students, both day school and evening. About one week later, 70 of these participated in the experiment.

RESULTS

Statistical properties of the several measures are given in Table 1. The correlations between the frequency of yielding and each of the four scales are presented as part of Table 2. Negro, Jewish, and Asian Ss were not included in correlations involving the E scale. IC, RI, and E showed slight but significant ($P < .025$, one-tailed test) correlations with frequency of yielding. The correlation with the F scale was .48 ($P < .01$), somewhat higher than Crutchfield's (1955) finding which it

³ This version of the F scale included the following items: Item 9, Form 78 (Adorno et al., 1950, p. 226); Item 44, Form 60 (Adorno et al., 1950, p. 248); Items 4, 6, 8, 9, 13, 18, 19, 25, 31, 35, 39, and 43, Forms 45 and 40 (Adorno et al., 1950, pp. 255-257).

replicates. The correlations of yielding with IC and RI were significantly lower than that with F ($P < .05$ in both cases). The correlation with the E scale was not significantly different from the correlation with F, the removal of cases required by this comparison causing the correlation with F to drop.⁴

The intercorrelations among the scales are also presented in Table 2. These show moderate to substantial positive relationships ($P < .01$) among the various measures. As a check against spurious correlation between IC and RI due to similarity of item content, two psychologists sorted the otherwise unidentified items into the proper ideological categories. Both judges correctly divided the items without an error.

DISCUSSION

It will be recalled that the F scale used in this study deliberately omitted items with conforming or rugged individualistic content. The remaining items express largely superstition, projectivity, sex preoccupation, moralism, and destructiveness. Summed up, these imply a weakening of the adaptive functions and lack of confidence in them. The higher correlation of yielding with the F scale than with IC and RI point to the conclusion that central personality variables as revealed by the F scale were more important determinants of yielding than more peripheral ideological variables as revealed by IC and RI. The evidence was equivocal

⁴ Comparison of r_{12} and r_{13} requires that $N = N_1 = N_2$. In the present instance, certain cases were dropped, and some of the correlations in Table 2 were recalculated. (See McNemar, 1949, p. 124.)

for the E scale, since the drop in correlation occasioned by removal of cases may have been due to the resulting homogeneity of variance. The authors of the F scale seem to have succeeded in their purpose of building an instrument that would measure those underlying personality tendencies responsible for antidemocratic ideology.

The more peripheral nature of the ideologies of conformity and rugged individualism probably accounts for the character of the relationship between them, low positive. Both are authoritarian ideologies, each having anti-individual as well as antisocial qualities. Yet, if they were *simply* dual and interchangeable expressions of authoritarianism, the correlation between them would be as high as their separate correlations with authoritarianism. Since this is not the case, it appears that they function only partly as expressions of authoritarianism, but partly as psychological alternatives. It might be rewarding to study the way these alternatives are used by yielders.

REFERENCES

- ADORNO, T. W., FRENKEL-BRUNSWIK, ELSE, LEVINSON, D. J., & SANFORD, R. N. *The authoritarian personality*. New York: Harper, 1950.
- ASCH, S. E. *Social psychology*. New York: Prentice-Hall, 1952.
- CRUTCHFIELD, R. S. Conformity and character. *Amer. Psychologist*, 1955, **10**, 191-198.
- MCNEMAR, Q. *Psychological statistics*. New York: Wiley, 1949.
- NADLER, E. B. The ideological correlates of conformity. Unpublished doctoral dissertation, Western Reserve Univer., 1957.

Received January 20, 1958.



OPUS

#1

**An event of importance for
NEW AUTHORS ONLY**

OPUS #1 is the title of a new Philosophical Library 1959 book series which will publish the deserving manuscripts of new authors only.

OPUS #1 has two basic requirements: The author must have something to say and know how to say it.

OPUS #1 will be dedicated to the humanities and will include works in the fields of Art, Literature, Philosophy, Religion, Sociology, History, Psychology and Psychiatry.

OPUS #1 books will be printed on special, fine antique paper, and artistically cloth-bound for library use.

Only complete book-length manuscripts accepted.

*Inquire for full details
Philosophical Library, Publishers
Dept. OP-105
15 East 40 St., N.Y. 16, N.Y.*

Schizophrenia Manfred Sakel	5.00
Epilepsy Manfred Sakel	5.00
The Analysis of Dreams Medard Boss	6.00
Group Psychoanalysis B. Bohdan Wassell	3.75
The Neuroses and their Treatment Edward Podolsky	10.00
Psychotherapy and Society Wladimir Eliasberg	6.00
The New Chemotherapy in Mental Illness H. L. Gordon	12.00
Principles of Self-Damage Edmund Bergler	6.00
The New Psychiatry Nathan Masor	3.75
Experimental Psychology Ivan Pavlov	7.50

Philosophical Library, Publishers
15 East 40 Street, New York 16

THE AMERICAN JOURNAL OF PSYCHOLOGY

Founded in 1887, The American Journal of Psychology is largely devoted to the interests experimental psychology, although its pages are open to workers in all fields of scientific psychology. The journal is published quarterly in March, June, September, and December. Subscriptions: \$7.00 a year (single numbers \$1.85); foreign, \$7.50 a year (single numbers \$2.00). Subscriptions, back numbers and volumes may be obtained from the Business Office.

PORTRAITS OF PSYCHOLOGISTS

Offprints of Portraits of Psychologists that have been published in The American Journal of Psychology are now available. Listed in order of their appearance, they are as follows:

1. G. Stanley Hall, Vol. 35, 1924, facing p. 313
2. Edmund C. Sanford, Vol. 36, 1925, facing p. 157
3. James Ward, Vol. 36, 1925, facing p. 449
4. Franz Brentano, Vol. 36, 1925, facing p. 469
5. Edward B. Titchener, Vol. 38, 1927, facing p. 489
6. Margaret Floy Washburn, Vol. 39, 1927, frontispiece
7. Herman Ebbinghaus, Vol. 42, 1930, facing p. 505
8. Wilhelm Wundt, Vol. 44, 1932, facing p. 615
9. Howard C. Warren, Vol. 46, 1934, facing p. 189
10. G. E. Müller, Vol. 47, 1935, facing p. 187
11. Carl Stumpf, Vol. 49, 1937, facing p. 173
12. Kurt Koffka, Vol. 55, 1942, facing p. 157
13. William James, Vol. 55, 1942, facing p. 309
14. Raymond Dodge, Vol. 55, 1942, facing p. 467
15. John A. McGeoch, Vol. 56, 1943, facing p. 1
16. James McKeen Cattell, Vol. 57, 1944, facing p. 133
17. Max Wertheimer, Vol. 57, 1944, facing p. 291
18. Joseph Jastrow, Vol. 57, 1944, facing p. 451
19. Charles S. Myers, Vol. 60, 1947, facing p. 159
20. Pierre Janet, Vol. 60, 1947, facing p. 479
21. James R. Angell, Vol. 62, 1948, facing p. 315
22. Knight Dunlap, Vol. 63, 1950, facing p. 1
23. Edward L. Thorndike, Vol. 63, 1950, facing p. 16
24. Carl E. Seashore, Vol. 63, 1950, facing p. 317
25. Oswald Külpe, Vol. 64, 1951, facing p. 1
26. Clark L. Hull, Vol. 65, 1952, facing p. 497
27. David Katz, Vol. 66, 1953, facing p. 519
28. Walter S. Hunter, Vol. 67, 1954, facing p. 571
29. Harvey A. Carr, Vol. 68, 1955, facing p. 1
30. Louis L. Thurstone, Vol. 69, 1956, facing p. 1
31. Madison Bentley, Vol. 69, 1956, facing p. 169
32. Robert M. Yerkes, Vol. 69, 1956, facing p. 345
33. Lewis M. Terman, Vol. 70, 1957, facing p. 333
34. Karl M. Dallenbach, Vol. 71, 1958, facing p. 1
35. Karl S. Lashley, Vol. 72, facing p. 1
36. John B. Watson, Vol. 72, facing p. 163

Separates—35¢ each
(Please order by number)

The series—\$9.00

THE AMERICAN JOURNAL OF PSYCHOLOGY

University of Texas Mezes Hall 102

Austin 12, Texas